Indian Journal of Animal Sciences **96** (4): 286–290, April 2026/Article
https://doi.org/10.56093/ijans.v96i4.165466

Comparative study of artificial neural network (ANN) and regression methods for prediction of breeding efficiency in Hariana cattle

RAKSHIT[1], VIJAY KUMAR[2]✉, S P SINGH[3], JITENDRA KUMAR AGRAWAL[4] and AMIT SINGH[5]

U. P. Pandit Deen Dayal Upadhyaya Pashu Chikitsa Vigyan Vishwavidyalaya Evam Go Anusandhan Sansthan (DUVASU), Mathura Uttar Pradesh 281 001 India

Received: 01 March 2025; Accepted:30 June 2026

ABSTRACT

The present investigation was undertaken to predict the breeding efficiency of Hariana cattle using Tomar method (BET) and early life reproduction traits. Under this study, both multiple linear regression (MLR) and artificial neural network (ANN) approach were used. Effectiveness of both methods was also compared for prediction of BE in Hariana cattle. Data of breeding information were collected from the history sheet registers and herd inventory registers of Hariana cattle maintained at DUVASU farm, Mathura. Data on productive animals born between 1962 and 2024 were recorded out of which 743 animal records were collected. Breeding efficiency was predicted by multiple linear regression analysis. In MLR study, it was observed that model 5 having two independent variables (AFC and FSP) fulfilled most criteria such as higher R^2, and lower MSE, RMSE, MAE, MAPE, and U values. In the present investigation, the accuracy of prediction obtained from ANN was more as compared to MLR for prediction of BET using early life reproduction traits in Hariana cattle. Multiple regression analysis was performed by considering different first lactation reproduction traits as independent variables and the breeding efficiency as dependent variables. The optimal model for BET included age at first calving and first service period. The optimal models were BET = 106.29 – 0.0192AFC – 0.029FSP. The MLR explained 69.6% of accuracy of prediction of BET in Hariana cattle with MSE of 42.20. With the ANN algorithm, the MSE was found to be 33.3, which is more precise as compared to MLR. For the prediction of BET, MLR method is simple and easy compared to ANN, but accuracy of ANN was more.

Keywords: ANN, Breeding Efficiency, MAE, MLR, RMSE

The economy of India is being strengthened by livestock. The cattle were divided into several breeds, with the high yielders being distinguished from the low or medium yielders. The Hariana cattle breed is a well-known native breed from Haryana, India, predominantly found in the districts of Rohtak, Hisar, Jind, and Gurgaon. These cattle are characterized by their white or light grey coat, sturdy and muscular physique, long legs, and a small hump. Primarily used as draught animals in agriculture, they also provide a moderate amount of milk, with cows producing approximately 5-8 liters per day under proper care. Renowned for their resilience, disease resistance, and ability to thrive in hot climates, this breed is highly valuable to farmers in northern India.

Present address: [1]PhD Scholar,[2]Associate Professor, [3]Professor, Department of Animal Genetics and Breeding, [4]Assistant Professor, Department of Veterinary Gynaecology and Obstetrics, [5] Professor, Department of Veterinary and Animal Husbandry Extension Education, College of Veterinary Science and Animal Husbandry, U. P. Pandit Deen Dayal Upadhyaya Pashu Chikitsa Vigyan Vishwavidyalaya Evam Go Anusandhan Sansthan (DUVASU), Mathura-281 001, Uttar Pradesh, India. ✉Corresponding author email: jaisvet@gmail.com

The primary goal of breeders has always been to assess the production potential of dairy animals. The productive and reproductive abilities of dairy animals are the foundation of the dairy industry's economics; thus, it makes sense to identify methods for enhancing these capacities through the development of specific selection criteria. According to Bakir and Cilek (2009), cattle reproduction is the first step in producing economic returns like milk, so the breeding efficiency of cattle reflects the profitable returns from the animal. Negussie *et al.* (1998) reported that the development of suitable breeding strategies requires an accurate assessment of the reproductive efficiency of native stocks in various situations of production. As the breeding efficiency is the lifetime trait of cattle, so selection of cattle on the basis of breeding efficiency of animal is too late as animal is aged at time of selection. Expenditure in maintaining animals, having less lifetime production will be reduced if selection done by early prediction of the phenotypic value of breeding efficiency from early reproduction traits. This further helps in culling process of less profitable animals in farms. According to Shetkar *et al.* (2021) one of the key factors in determining the maximum financial return from a specific animal is lifetime

productivity. In light of the aforementioned information, research was conducted on Hariana cattle with the aim of creating a predictive model for breeding efficiency by multiple regression analysis and artificial neural network.

MATERIALS AND METHODS

Source of data: Breeding information were used from the history sheet registers and herd inventory registers of Hariana cattle maintained at DUVASU farm, Mathura. Data on productive animals born between 1962 and 2024 were gathered. Total 743 animal records were collected from Farm to predict breeding efficiency. To ensure the normal distribution, the outliers were removed and data within the range of mean ± 3SD were considered for the study.

Breeding efficiency was estimated by formula given by Tomar (1965).

$$BET = \frac{[(N-1)\ 365 + 1020]}{AFC + D} \times 100$$

Where,
BE - Breeding efficiency in percentage
N - Number of calvings
AFC - Age at first calving in days
D - Number of days from first to last calving

Means, standard deviations, standard errors and coefficients of variations of the BET were calculated according to the statistical method given by Snedecor and Cochran (1994).

To measure and quantify how much the variance is inflated VIF tool is used. When correlation exists among predictors the value of coefficient of determination is biased. In that case standard error of predictors coefficients will increase and consequently the variance of predictors coefficients is inflated (Kumar *et al.* 2019). VIF can be calculated using the formula:

$$VIF = \frac{1}{1-R_i^2}$$

Where R_i^2 is the R^2– value obtained by regressing the i^{th} predictor on the remaining predictors. A variance inflation factor exists for each of the i predictors in a multiple regression model.

Multiple regression analysis: The multiple regression analysis was done as suggested by Draper and Smith (1987) with the help of following model:

$$Y_i = a + b_1X_1 + b_2X_2 + \ldots. + b_nX_n + e_i$$

Where,
Y_i - Variable to be predicted
a, b_1, b_2, …, b_n - Unknown parameters to be estimated.
X_1, X_2, …, X_n - Traits whose values are known.
e_i - Random error, NID (0, σ_e^2)

Estimation of coefficient of determination (R^2): Out of hundred per cent of variability of the prediction how much variation was contributed by a set of independent variables and is expressed in terms of percentage is called the coefficient of determination (R^2). The R^2 estimated in different models from analysis of variance:

Development and selection of optimum model for prediction of breeding efficiency in Hariana cattle: R^2, mean absolute error (MAE), mean absolute percentage error (MAPE), Theil's U-statistics (U), and root mean squares error (RMSE) values were used to select the optimal model or models for predicting BET in Hariana cattle. The best models were those with greater coefficients of determination (R^2) and lower RMSE, MAE, MAPE, and RMSE values. Additionally, models generated using artificial neural networks (ANN) were compared with the best model of multiple regression analysis (MLR). According to Kumar *et al.* (2019) to predict the response variable using the predictor variables is the main objectives of model development. It is practically significant to evaluate a model's prediction ability because model development for prediction takes a lot of research work as well as time consuming. Hence, the model developed should be simple and have good prediction percentage.

Artificial neural network (ANN): A multilayer feed forward neural network with a back propagation error learning method was used in this research, designed for forecasting breeding efficiency by Tomar method (BET) using the R programming language (R core team, 2013). In order to train and simulate the ANN, resilient back propagation methods were used. To achieve reliable findings, the network was trained using a training data set. A fresh dataset (test data) was used to evaluate the prediction capability. When given fresh inputs, a well-trained ANN is likely to provide output that makes sense. The entire collection of data was randomly split into two subsets: the testing subset, which included 33.7%, 25%, or 15% of the data, and the training set, which included 66.3%, 75%, or 85% of the data. Neural network models were trained using the training sets, and the models were validated using the testing sets. Two hidden layers (10:5, 2:1 neurons) and were used to test the network. MLR, as reported by Draper and Smith (1987) and ANN were used to predict BET using early lactation traits. The mean square errors of the two methods were used to compare the effectiveness of each method.

RESULTS AND DISCUSSION

Measure of central tendency and measure of dispersion: Total 743 animal data were collected for breeding efficiency. After normalization, one outlier was removed to estimate overall mean for breeding efficiency. BET was 63.9% with standard error and standard deviation of 0.44 and 11.8 respectively. The coefficient of variation estimated for breeding efficiency was 18.52. The higher mean of breeding efficiency was reported by Habib *et al.* (2013) in Red Chittagong Cattle and by Ebrahim (2018) in Friesian cattle.

Development of MLR model for breeding efficiency: To determine the correlations between the independent traits in the study of prediction, the variance inflation factor (VIF) was employed. The lower VIF values for AFC, FSP and FDP were observed which means that these traits were

not correlated. Multiple regression analysis was carried out for the prediction of BET. To predict BET the early reproduction traits like AFC, FSP and FDP were used as independent variables for multiple linear regression method. Step wise regression procedure was used to predict the breeding efficiency from first lactation traits, viz. AFC, FSP and FDP under the study. On analysis the co-efficient of determinations of 47.4, 15.99, 26.65, 61.81, 69.66, 26.70 and 69.83% for BET were observed when AFC, FDP, FSP, AFC-FDP, AFC-FSP, FDP-FSP and AFC-FDP-FSP were used as independent variables, respectively. The models for prediction of BET using early production and reproduction traits developed are presented in the Table 1. The developed models were in different combination of AFC, FSP and FDP traits. Out of 07 models developed, model number 05 having two independent variables i.e., AFC and FSP has fulfilled many of the criterions such as higher R^2 (co-efficient of determination) and lower RMSE, MAE, MAPE, and Theil's U-statistic value values than the rest of the models. Different criterion values considered for judging the optimum model for BE in Hariana cattle are given in Table 1. The correlation coefficient between the actual and predicted values of best model developed using MLR for BET was 0.83.

The linear regression equation developed is given below.

$$BE\ (T) = 106.29 – 0.0192AFC – 0.029FSP$$

The estimated co-efficient of determination from the model was 69.6%. The R^2 value of prediction was relatively high suggesting that the relationship between the predictors and the response variable is linear. Tomar *et al.* (2023) predicted the BET with lower co-efficient of determination of 57.9%. Shetkar *et al.* (2023) predicted the lifetime 2 milk yield with higher co-efficient of determination of 75.46%. Singh *et al.* (2021) used MLR to predict the first lactation 305-day milk yield (FL305DMY) in Murrah buffalo by using stepwise backward regression analysis. For the prediction of FL305DMY best combination of monthly test-day milk yields were selected based on adjusted R^2 and RMSE values. The prediction of 305-day milk yield from monthly test-day milk yields using TD-2 (45th day), TD-4 (105th day) and TD-6 (165th day) combination gave higher accuracy (82.76%) of prediction.

Development of ANN model: The ANN was trained using the training data set comprising the variables AFC, FSP, and FDP, which were included in the regression analysis optimal equation. The network was trained using a variety of hidden layer combinations with different numbers of neurons; the combination of two hidden layers produced the best results. MSE values with nodes considered of different ANN model (SET-A, SET-B and SET-C). In comparison to SET-A (Training-test data 66.3:33.4) and SET-C (Training-test data 85:15), set-B (Training-test data 75:25) had the lower MSEs, 35.51 (for node 10:5) and 33.38 (for node 2:1). Fig. 1 show the plotted regression of the expected BET on the actual BET predicted by ANN. The correlation coefficient between the actual and predicted values of best model developed using ANN for BET was 0.85. Mundhe *et al.* (2014) reported a higher R^2 (89.29%) for an ANN when predicting FL305DMY using monthly part lactation milk yield records in Sahiwal cattle. Bhosle and Singh (2017) develop lifetime milk yield (LTMY) prediction equation using different economical traits in Holstein Friesian crossbred dairy cattle in India. The BR neural network model was able to predict milk yield with 71.18% R^2. Singh *et al.* (2020) used ANN for prediction of FL305DMY in Murrah buffaloes. They trained three algorithms that were Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG). Each algorithm was compared using four different training-test data sets (66.66:33.33, 75:25, 80:20 and 90:10). BR achieved highest accuracy of 81.98% with lowest RMSE value of 16.13 % for input set-1 and 78.33% accuracy with RMSE value of 16.89 % for input set-2.

Comparison between the best model of MLR and ANN: With the help of multiple linear regression analysis, BET can be predicted with R^2 of 69.6% using age at first calving and first service period in Hariana cattle. When ANN was used in the current study to predict BET using early reproduction traits in Hariana cattle, the prediction accuracy was nearly identical to that of MLR. Although the accuracy of ANN is more accurate, the MLR technique should be chosen because it is simpler than ANN. Fig. 1 graphically represent the BET predictions produced by the best ANN model. The MLR explained 69.6% of accuracy of prediction of BET in Hariana cattle. The best ANN

Table 1. Prediction equation for breeding efficiency (Tomar)

Sl. No.	Dependent variable	Independent variables	R^2 (%)	Linear regression equation's	RMSE	MAE	MAPE	U	MSE
1	Breeding efficiency (Tomar)	AFC	47.4	BE(T) = 98.17 -0.0201AFC	8.55	7.06	11.44	0.13	73.10
2		FDP	15.9	BE(T) = 72.11 - 0.025FDP	10.80	8.78	14.17	0.64	116.64
3		FSP	26.6	BE(T) = 75.037 – 0.032FSP	10.09	8.12	13.14	0.15	101.81
4		AFC, FDP	61.8	BE(T) = 104.86 - 0.0198AFC – 0.024FDP	7.28	5.77	9.13	0.11	52.99
5		AFC, FSP	69.6	BE(T) = 106.29 – 0.0192AFC – 0.029FSP	6.49	5.12	8.20	0.09	42.12
6		FDP, FSP	26.7	BE(T) = 75.16 – 0.031FDP - 0.002FSP	10.09	8.11	13.08	0.15	101.81
7		AFC, FDP, FSP	69.8	BE(T) = 106.598 – 0.019AFC – 0.027FSP – 0.004FDP	6.48	5.1	8.18	0.09	41.99

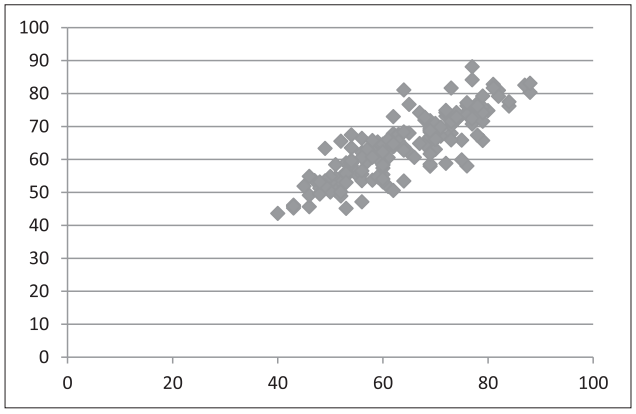

Fig. 1. Actual versus predicted BET for the best ANN model in Hariana cattle

algorithm accuracy of prediction for optimum model has mean sum of square error (MSE) 33.3 which was low as compared to the MLR model for BET with MSE 42.20. It was found that the correlation coefficient between the two approaches predicted outcomes was 0.98. A high correlation obtain indicates that the two techniques produced results that were equivalent. Although the accuracy of ANN is more, the MLR technique can be chosen because it is simpler than ANN. MLR and ANN showed relatively similar predictive performance in the present investigation because the relationship between breeding efficiency by Tomar method (BET) and the selected early-life reproductive traits appears to be largely systematic and moderately predictable, with limited nonlinear complexity. Several factors may explain this observation viz. Strong influence of a few key reproductive traits, predominantly linear relationship among variables, limited number of predictor variables and breeding efficiency may have biological constraints.

Usman *et al.* (2025) in their study indicate that FL305DMY can be predicted in Vrindavani cows early in the first lactation using the first four test-day milk yields and the first four monthly milk yields, along with the AFC and PY, using an ANN. In their study, the R^2 value was higher for the ANN than the MLR, and the RMSE values were lower for the ANN than the MLR. Bhosale and Singh (2015) reported that an ANN model had a slightly higher accuracy (85.07%) than an MLR model (84.60%) in predicting FL305DMY using test-day milk records. Subhita *et al.* (2023) inferred from their study that artificial neural network was better than the multiple linear regression to predict FL305DMY with more than 70% accuracy with lesser value of RMSE in comparison to MLR for the prediction of the first lactation 305 days milk yield in Tharparkar cattle. The higher accuracy obtained in these studies may have been because of their use of more input variables.

In the present study, BET was predicted by multiple regression analysis (MRA). The optimum models were selected based on their higher co-efficient of determination (R^2) and lower root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), Theil's U-statistic values. The present study demonstrated the potential of early-life reproductive traits for reliable prediction of breeding efficiency by Tomar method (BET) in Hariana cattle. Among the multiple linear regression (MLR) models evaluated, the model incorporating age at first calving (AFC) and first service period (FSP) emerged as the most effective, indicating that these early reproductive indicators can serve as practical selection criteria for improving breeding performance. However, the artificial neural network (ANN) approach exhibited superior predictive accuracy over MLR, as reflected by its lower mean square error, highlighting its greater capability to capture complex relationships among reproductive variables.

These findings have broader implications for livestock breeding and herd management by demonstrating that breeding efficiency can be estimated at an earlier stage of an animal's productive life rather than waiting for complete lifetime reproductive records. Such predictive approaches can support timely selection decisions, accelerate genetic improvement, reduce generation intervals, and enhance overall herd productivity. From a practical perspective, MLR offers a simple, interpretable, and easily implementable tool for routine farm-level applications, whereas ANN provides a more advanced and precise alternative for data-driven breeding programs and precision livestock management systems.

The study further emphasizes the growing relevance of integrating machine learning approaches into animal breeding research. The adoption of predictive models such as ANN in organized breeding farms and decision-support systems could contribute to more efficient resource utilization, improved reproductive performance, and sustainable genetic improvement strategies in indigenous cattle populations such as Hariana cattle.

REFERENCES

Bakir G and Cilek S. 2009. A research on reproductive traits of holstein cattle reared at tahirova state farm in Balikesir province in Turkey. *Journal of Animal Veterinary Advances* **11**(8): 2383–2387.

Bhosale M D and Singh T P. 2015. Development of lifetime milk yield equation using artificial neural network in Holstein Friesian crossbred dairy cattle and comparison with multiple linear regression model. *Current Science* **108**(12): 2257–2260.

Bhosale M D and Singh T P. 2017. Development of lifetime milk yield equation using artificial neural network in Holstein Friesian crossbred dairy cattle and comparison with multiple linear regression model. *Current Science* **113**(5): 951–955.

Draper N R and Smith H. 1987. *Applied Regression Analysis*. John Wiley and Sons Inc., New York.

Ebrahim S. 2018. Genetic and non genetic factors affecting the breeding efficiency in Friesian Cows in Egypt. *Journal of Animal and Poultry Production* **9**(12): 459–461.

Habib M S, Bhuiyan A, Amin M R and Khan M A S. 2013. Seasonal index and breeding efficiency of red Chittagong cattle in Bangladesh. *Journal of Tropical Resources and Sustainable Science* **1**: 55–61.

Kumar V, Chakravarty A K, Magotra A, Patil C S and Shivahare 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): 1262–1268.

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): 517–518.

Negussie E, Brannang E, Banjaw K and Rottmann O U. 1998. Reproductive performance of dairy cattle at Assella livestock farm. Arsi. Ethiopia. I: Indigenous Cows Versus their F1 Crosses. *Journal of Animal Breeding and Genetics* **115**: 267–280.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Shetkar M, Kumar V, Singh S P, Singh Y Kumar M and Nath S. 2021. Age at first calving and lifetime performance of Hariana cattle at organized farms. *Indian Journal of Animal Sciences* **91**(12): 1106–1108.

Shetkar M, Kumar V, Singh S P, Singh Y and Kumar M. 2023. Modeling and forecasting of lifetime milk production in Hariana cattle. *Indian Journal of Animal Sciences* **93**(4): 406–409.

Singh M, Prakash V, Das S, Dixit S and Gupta A K. 2021 Multiple linear regression analysis using monthly test day milk yield predicting the first lactation production performance for sire evaluation in Murrah buffaloes. *Indian Journal of Dairy Science* **74**(1): 89–94.

Singh N, Usman S, Maurya V, Dutt T, Bhatt N, and Kumar A. 2020. Comparative analysis of artificial neural network algorithms for prediction of FL305DMY in Murrah buffalo. *International Journal of Livestock Research* **10**(9): 205–209.

Snedecor G W and Cochran W G. 1994. *Statistical methods (8th edition)*. Calcutta, India: Oxford and IBH Publishing.

Subhita, Nehara M, Pannu U, Bairwa M and Rashmi. 2023. Comparative study of multiple linear regression and artificial neural network for prediction of first lactation 305-days milk yield in Tharparkar cattle. *Indian Journal of Dairy Science* **76**(1): 64–68.

Tomar K, Kumar R, Tyagi K K, Sahu D S, Yadav S P, Gupta A, Singh D and Kumar D. 2023. Prediction of lifetime performance traits from first production and reproduction traits in Hariana cattle. *Journal of Animal Research* **13**: 739–744.

Tomar N S. 1965. A note on the method of working out breeding efficiency in Zebu cows. *Indian Dairyman* **17**: 389–390.

Usman S M, Dutt T, Sahib Q S, Singh N P, Tiwari R, Chandrakar J, Abo Ghanima M M, Youssef I M, Sherasiya A, Kumar A and Swelum A A. 2025. Artificial neural networks for predicting first-lactation 305-day milk yield in crossbred cattle. *South African Journal of Animal Science* **55**: 1–9