SUPPLEMENTARY MATERIAL AVAILABLE ONLINE

Indian Journal of Animal Sciences **96** (5): 404–409, May 2026/Article
https://doi.org/10.56093/ijans.v96i5.179224

Decision support system for selection of dairy buffalo based on peak yield

EKTA HOODA[1], NARESH KUMAR DAHIYA[2], ASHOK KUMAR BALHARA[1], ASHOK BOORA[1], SARITA YADAV[1], S K PHULIA[1] and SUNESH BALHARA[1✉]

ICAR-Central Institute for Research on Buffaloes, Hisar Haryana 125 001 India

Received: 20 May 2026; Accepted: 29 May 2026

ABSTRACT

Efficient animal selection is vital for optimizing dairy farming. Traditionally, human reliance introduces errors. In the present study, a user-friendly Decision Support System (DSS) using Artificial Neural Networks (ANNs) was developed to predict peak milk yield of buffalo (Murrah breed) based on linear traits - Udder Depth (UD), Naval Udder Distance (NUD), Rear Teat Distance (RTD), Fore Rear Teat Distance (FRTD), Teat Length (TL), Rump Width (RW), Rear Udder Width (RUW), Rear Udder Height (RUH), and Lactation Number (LN). A dataset of 259 buffaloes was used for model training and validation, with ANN outperforming Support Vector Machine, Random Forest and Multiple Linear Regression. The performance of ANN was assessed with the Correlation Coefficient (CC), Root Mean Squared Error (RMSE) and R^2 metrics, revealing maximum CC of 0.86, minimal RMSE of 2.03 and a maximum R^2 of 0.748, highlighting its accuracy. These findings provide valuable insights for dairy farmers, enabling informed decisions in selecting elite buffaloes. Implementation of this tool has the potential to revolutionize breeding, boost milk production, and enhance overall efficiency and profitability in buffalo based dairy farming.

Keywords: Animal selection, ANN, Linear traits, Machine learning

Indian agriculture is currently grappling with a pressing issue, the profitability of the industry. The struggles once tied to crop agriculture have now shifted towards livestock, where the need is to improve efficiency with limited resources. In India, buffaloes remain the preferred dairy animal because of their better feed conversion efficiency, tolerance to harsh climate and long productive life (Balhara *et al.* 2017).

The dairy sector has shown a steady rise in milk production over the years, largely due to selective breeding (Sharma *et al.* 2024). Selection of high-producing buffaloes through linear traits has so far depended mostly on traditional knowledge. As per the "Guidelines for Type Classification of Cattle and Buffalo" (NDDB, 2017), linear traits refer to specific characteristics that define an animal's body conformation. An animal's physique gives useful clues about its health, milk yield, and reproductive performance (Sahu *et al*. 2024, Dahiya *et al.* 2020). Linear type classification therefore plays an important role in herd-level decisions, helping identify animals with higher productive potential.

But human judgement in this area is often error-prone, especially when the evaluator is not fully aware of desirable dairy characteristics. A wrong call-in buffalo selection affects return directly, and also through the offspring of the selected animals. As an alternative to human experts, AI-based automated decision platforms are now being seen as a major shift in dairy farming, since they can learn from past data and make decisions in a way close to human reasoning (Slob *et al.* 2021). Among soft computing tools, artificial neural networks (ANN) are widely used for predicting milk production potential in dairy animals (Grzesiak *et al*. 2025). In buffaloes, predictive modelling has also been applied to traits such as physiological assessment using infrared thermography (Balhara *et al*. 2024) and estrus detection through urinary metabolite signatures (Sangwan *et al.* 2025).

Despite these advances, little work has been done on ANN-based Decision Support Systems (DSS) for buffalo selection. Keeping this gap in mind, the present study was planned to develop an ANN-based DSS for selecting dairy buffaloes on the basis of their linear traits. The hypothesis was that such a system can improve selection accuracy and herd productivity compared to conventional methods.

MATERIALS AND METHODS

Animal selection and data acquisition: This investigation made use of a previously used dataset, created by the same team of researchers (Balhara *et al*. 2022), to construct a predictive model-based DSS. A total of 259 lactating buffaloes from the Murrah breed in their first to fifth parity were selected for this study. The complete data collection methodology, variable definitions and preprocessing steps

Present address: [1] ICAR-Central Institute for Research on Buffaloes, Hisar- 125 001, India.[2] ICAR- Division of Agricultural Education, KAB - II, New Delhi - 110 012, India. ✉Corresponding author email: balharas@gmail.com

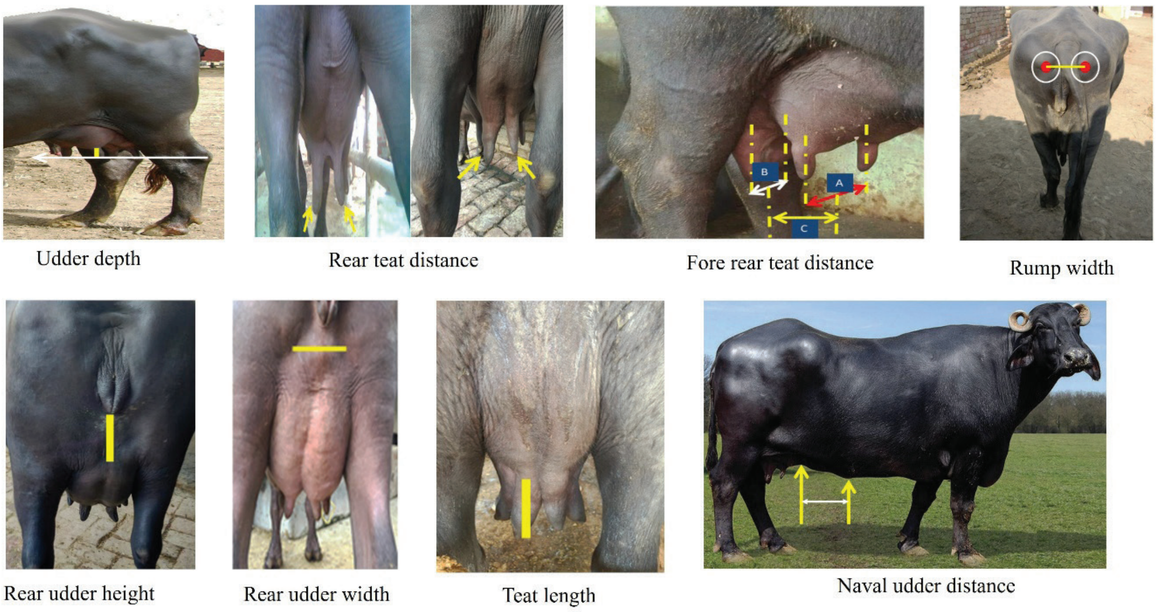

Fig. 1. Measurement of the linear morphological traits (UD, RTD, FRTD, RW, RUH, RUW, TL and NUD) used for predicting peak milk yield in Murrah buffaloes. Yellow/white markers indicate the reference points and direction of each measurement

such as data cleaning, normalisation and feature selection were comprehensively done as described by Balhara *et al.*2022. Initially fourteen input variables were considered, out of which nine variables, comprising eight linear morphological traits and lactation number, were selected based on their predictive importance for peak yield.

These selected variables included Udder Depth (UD), Naval Udder Distance (NUD), Rear Teat Distance (RTD), Fore Rear Teat Distance (FRTD), Teat Length (TL), Rump Width (RW), Rear Udder Width (RUW), Rear Udder Height (RUH), and Lactation Number (LN) (Figure 1). Consequently, this optimal subset of eight distinctive linear traits, in addition to the lactation number, was ultimately utilized in developing the DSS.

Design and development of DSS: Following the conclusions of Balhara *et al*. 2022, a multi-layer feed-forward neural network was employed, using a back-propagation learning algorithm to develop a DSS for predicting peak milk yield. Among the four previously tested models-Multiple Linear Regression (MLR), Support Vector Machine (SVM), Random Forest (RF) and ANN-the ANN model, as reported by Balhara *et al*. 2022, showed the best performance with lowest RMSE (2.03) and a maximum R^2 of 0.748 and CC of 0.86.

Several key parameters, including the number of layers, the number of neurons within each layer, learning rate, momentum, and others, played a pivotal role in training the model. During experiments, these parameter values were fine-tuned to achieve optimal performance. To minimize bias and enhance accuracy, the ANN models were trained and validated using a 10-fold cross-validation method. This approach involved treating each dataset as both training and testing data. CC, RMSE and R^2 were used as evaluation criteria for comparison of models.

The optimal architecture for the ANN was determined through a trial-and-error approach, involving variations in the number of hidden layers from 1 to 3 and neurons on each hidden layer from 2 to 10. Experiments were also conducted with different learning rates and momentum values (0.01, 0.05, 0.1, 0.2, 0.4, and 0.6). The sigmoid activation function was used for non-linearity, except at the output node. Training was carried out over epochs of 100, 500, 1000, and 2000. It is acknowledged that with the current dataset of size n=259, heavy tuning of layers, neurons, learning rate and momentum could lead to overfitting and weaker generalization. In order to limit this, the number of inputs were reduced with the help of feature selection, choosing a small architecture over larger ones and a 10-fold cross validation.

Among the various configurations tested, the most favourable one consisted of 2 hidden layers, with 5 neurons in the first layer and 3 neurons in the second layer. This architecture yielded the highest CC of 0.8639, maximum R^2 of 0.74 and the lowest RMSE of 2.0308, establishing it as the optimal choice for constructing the proposed DSS. The structure of the optimal ANN model, obtained from WEKA (2023), is illustrated in Supplementary Figure 1, while its weights and biases are summarized in Tables 1, 2 and 3. These parameters enabled the model to capture the complex non-linear relationship between input variables and the predicted peak milk yield.

A trained ANN network can be converted into a mathematical formula through the weights and biases in combination with the transfer function (Lawal, 2020). The output Y of the ANN network can be expressed mathematically as in Equation (1):

$$Y=\sum_{(k=1)}^{3} W_{3_k} \cdot f_{sig}(b_{h_2k}+\sum_{(j=1)}^{5} W_{2jk} \cdot f_{sig}(b_{h1_j}+\sum_{(i=1)}^{9} W_{1ij} \cdot x_i))+b_0 \quad (1)$$

Here, x_i is the normalized value of the i^{th} input variable,are the weights from the input layer to the first hidden layer, are the weights from the first hidden layer to the second hidden layer and are the weights connecting the second hidden layer to the output neuron. Where , represent the bias terms at the respective layers. Y is the normalized output variable and is the sigmoid transfer function, defined as $f_{sig}(z)=1/(1+e^{-z})$.

In the mathematical model proposed for predicting ANN in this study, there are nine normalized input variables and one targeted output variable. The number of neurons connecting the input and the first hidden layer is five, and the number of neurons connecting the first and the second hidden layer is three. To achieve dimensionality consistency of the parameters and compatibility with the sigmoid transfer function, both the input and output parameters were normalized to fall within the range of -1 to 1, as described in Equation (2).

$$Input_{Norm} = 2* \frac{Input\text{-}Attribute_{Min}}{Attribute_{Max}\text{-}Attribute_{Min}} -1 \quad (2)$$

Where $Input_{Norm}$ is the normalized parameters, $Attribute_{max}$ the maximum value of the actual data, is the minimum value of the actual data, is the actual data to be normalized.

Since Y in Equation (1) is in normalized form, to get the actual Peak Yield (PY), the output of Equation (1) was de-normalized. The de-normalized form of Y is as presented in Equation (3)

$$Peak\ Yield_{denorm} = (Y+1) \times \frac{Attribute_{Max}\text{-}Attribute_{Min}}{2} + Attribute_{min} \quad (3)$$

Development of ANN-based DSS: To operationalize the ANN model, the ANN weights and biases from WEKA implementation were integrated into a python script using NumPy. To maintain the consistency with the model training process, the input values are automatically normalised using the same min-max scaling technique applied during model development. The same architecture was maintained and the normalised inputs were passed through two hidden layers using sigmoid activation functions. The extracted weights and bias values were fed into the program. The predicted output was subsequently deformalized to retrieve the actual peak milk yield value. The Graphical User Interface (GUI) was developed using the popular *Tkinter* library (2021) in Python to facilitate user interaction. The interface allowed users to input values of nine selected traits and obtains real time peak yield prediction. The interface included essential error handling functions to ensure all values were numeric and within expected biological limits, thus preventing invalid entries. This GUI effectively transformed the trained ANN model into a usable DSS for the field.

Table 3. Weight and bias at output layer

Output neuron	Weights			Bias
	B1	B2	B3	
C	-1.8294	-1.6725	-1.9022	1.0150

RESULTS AND DISCUSSION

The descriptive statistical analysis was performed for complete dataset (n=259, 15 attributes including one dependent attribute) using SPSS (version 20.0) for understanding the distribution and relationship among different traits. Result of this analysis in terms of range, mean, standard error, coefficient of variation and correlation with PY of each trait are detailed by Balhara *et al.* (2022). Key findings showed that peak milk yield had highly significant correlation with some of the input variables. Specifically, RUW (r = 0.572), LN (r = 0.567), and FRTD (r = 0.539) exhibited statistically significant

Table 1. Weights and biases at hidden layer 1

Hidden neurons	Weights									Biases
	UD	NUD	RTD	FRTD	TL	RW	RUW	RUH	LN	
A1	2.0596	0.5698	3.1806	-0.5663	1.3333	1.9451	2.9668	-2.2745	2.0662	1.4888
A2	-3.1945	-3.0598	1.4179	3.6100	-0.1542	1.9956	-1.3943	0.2998	0.9895	-0.9146
A3	0.6385	-0.7853	-1.0542	2.2091	1.9499	1.2102	1.3354	-2.8019	-2.3967	-3.6478
A4	-2.0432	-0.7494	1.8893	-0.8260	1.03423	-1.8010	0.3406	-0.3497	3.6728	-1.7516
A5	-1.0497	0.07397	1.1191	-0.3981	1.2013	-0.3238	1.8784	-1.0087	1.0045	-1.9746

Table 2. Weights and biases at hidden layer 2

Hidden Neurons	Weights					Biases
	A1	A2	A3	A4	A5	
B1	0.1942	-3.5063	-2.6186	-0.2077	2.0645	-1.667
B2	-0.4992	-0.9043	-1.6414	-1.7294	0.5837	-0.8222
B3	-1.0353	0.5311	-1.3159	-2.5651	0.07841	-0.1703

Table 4. Performance evaluation of predictive modelling approaches under consideration

Evaluation Matrices	ANN	SVM	RF	MLR
RMSE	2.03	2.17	2.12	2.15
R^2	0.748	0.706	0.723	0.711
Correlation Coefficient	0.86	0.84	0.85	0.84

positive linear relationships with PY, indicating that as these variables increase, PY tends to increase as well. On the other hand, RUH (r = -0.680) and NUD (r = -0.542) displayed significant negative linear relationships with PY, suggesting that an increase in RUH or NUD was associated with a decrease in PY. To address multicollinearity and enhance model stability, feature selection was performed using Cfs Subset Eval algorithm with Best First Search in WEKA. Following the key findings of Balhara *et al*. (2022), who developed and compared four predictive modelling approaches- MLR, SVM, RF and ANN, a summary of the comparative performance of these models is given in Table 4.

Among these, the ANN model emerged as the most accurate and consistent as indicated by the bold values given in the Table 4. It demonstrated superior capability in capturing complex, non-linear relationship between linear traits and peak milk yield. The best performance was obtained with the combination of 2 hidden layer with 5 neurons in first hidden layer and 3 neurons in second hidden layer in 1000 epochs with learning rate as 0.1 and momentum as 0.4. The CC and RMSE for the optimum architecture of ANN model for predicting peak yield were 0.86 and 2.03, respectively. Based on these results, a well-trained ANN model, having the highest performance criteria, could then be used to develop DSS to select a buffalo on the basis of peak milk yield. Supplementary Figure 2 shows scatter plot of comparison of actual and predicted peak yield using optimum ANN model.

Various studies suggested that ANN modelling could be used for 305 days milk yield forecasting (Dongre *et al*. 2012; Aruna *et al*. 2022; Usman *et al*. 2020) at early stage of lactation with high degree of accuracy than conventional method of prediction in different breeds of dairy cattle. A study (Singh *et al*. 2020) reported that an ANN model successfully predicted 305-days milk yield with an 82% accuracy using first lactation records in Murrah buffaloes. Similarly, Kumar *et al.* (2019) compared an ANN model with MLR for the prediction of first lactation milk yield based on test day records in Murrah buffaloes, yielding similar performance, both achieving a 76% accuracy. For the estimation of first lactation 305-day milk yield, the study found that MLR provided the most accurate results, followed by ANN and the Centering Date Method (CDM), using bimonthly test day milk yield data (Rana *et al.* 2021). ANNs have demonstrated their effectiveness in milk yield estimation, not only in cattle and buffalo but also in sheep and goats (Angeles-Hernandez *et al*. 2022, Fernández *et al.* 2007, Kaygisiz Sezgin 2017).

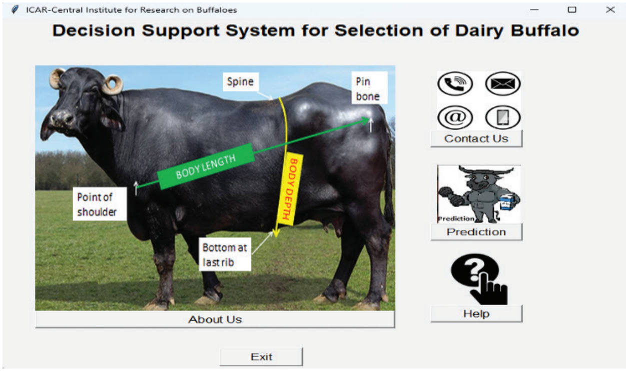

Fig. 2. Main Window of the Decision Support System (DSS)

The procedural flow for executing the DSS was described below: as depicted by visual representations in Figure 2, the main window of the DSS served as the first page of the system and consisted of five sections: Contact Us, About Us, Prediction, Help and Exit along with the title of the DSS, which indicated the purpose of the DSS. Upon clicking the "About Us" or "Contact Us" buttons, the GUI launched a new window for each respective section. Within these sections, users can explore detailed descriptions, gaining insights into the project's goals, contributors, and contact information. The "Help" button serves as a valuable resource for users seeking guidance or information about the GUI and its functionalities. The "Exit" button offers users a way to end their session with the GUI.

The section labelled "Prediction" in Figure 2 is a critical component of the application. On clicking on the "Prediction" section within the main window, an input window (Figure 3) opens, allowing input of animal tag numbers and nine input variables. Additionally, it provided options for initiating a new test, viewing the model output, and closing the interface. The "Predict" button in Figure 3, positioned below the input fields for parameters, initiated the prediction process. This process involves taking the user-provided input parameters and forwarding them to the underlying ANN model for processing and analysis. This intricate procedure encompassed a series of complex mathematical computations, including a crucial step of normalizing the input data to ensure it aligns with the model's specific requirements and scaling. Subsequently, the normalized data passes through the model's hidden layers of nodes, each equipped with unique weights and biases acquired during the model's training phase. Activation functions, such as sigmoid and linear functions, fine-tune the data as it traversed these layers, ultimately transforming it into a prediction. This prediction was then denormalized to provide a real-world interpretation. The result is presented to the user, offering valuable insights based on the input parameters.

The result window (output popup), depicted in Figure 4, promptly appears once the prediction process is complete, ensuring users receive the output without delay. The result screen displayed the expected peak milk yield, which is a prediction generated by our machine-learning model.

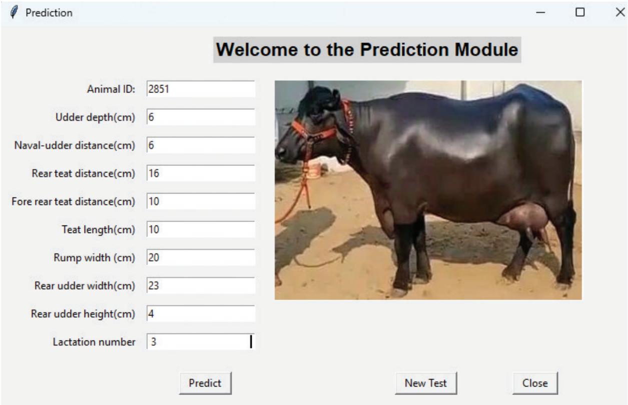

Fig. 3. Prediction Window of the Decision Support System (DSS) (Input Window)

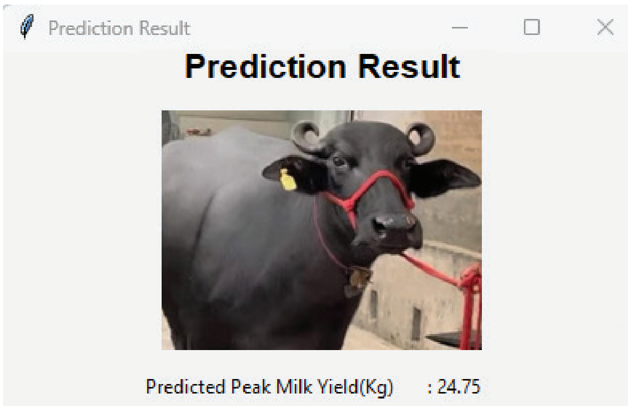

Fig. 4. Prediction Result Window of the Decision Support System (DSS) (Output Window)

To verify that python deployment functions identically, few randomly selected records from the existing data set were used as test inputs. The output from the GUI matched exactly with those predicted by the WEKA tool. This perfect alignment confirmed that the ANN logic, activation function and normalisation/denormalization processes were replicated successfully.

The DSS model's capability to accurately estimate peak milk yield, based on linear traits, provided valuable insights for buffalo breeders and farmers. It empowers them to make data-driven decisions regarding breeding choices, nutritional plans, and overall herd management practices. The precision of the model minimized guesswork and allowed for more efficient resource allocation, ultimately leading to increased milk production and herd quality. Thus, the DSS developed in this study demonstrated good potential for facilitating selection of dairy buffaloes by accurate prediction of peak milk.

Strengths and limitations: A key strength of this study is that it turns a validated ANN model into a simple, field-ready DSS, letting farmers get peak yield predictions from easily measured linear traits without any statistical background. Keeping the inputs less in number and a small network validated across 10-fold cross validation further adds to the tool's reliability.

The model suffers limitations too. The model was built on a fairly small, single breed (Murrah) dataset, and hence, its predictions may not carry over directly to other breeds, herds and management systems. It also depends only on linear traits and LN and does not consider other important factors that may affect peak yield such as season of calving or temperature-humidity index (THI). A further constraint is that little work is available on studies using linear traits of dairy buffalo for milk yield prediction. Therefore, it was not possible to compare the predictive performance of the models created in the current study with others. Comparisons between models is also difficult due to differences in data sets and variables used to create the model (Murphy *et al*. 2014). Future work needs to test the system on larger, multi-breed datasets and bring management and environmental variables to improve accuracy of the DSS.

Peak milk yield prediction is an important criterion in dairy animal selection. This study developed a DSS for buffaloes, validated through data-driven methods to ensure reliability and ease of use. Expanding the dataset will further improve prediction accuracy, while future work should focus on integrating image processing or computer vision techniques to reduce manual trait measurement and minimize errors.

REFERENCES

Angeles-Hernandez J C, Castro-Espinoza F A, Peláez-Acero A, Salinas-Martinez J A, Chay-Canul A J and Vargas-Bello-Pérez E. 2022. Estimation of milk yield based on udder measures of Pelibuey sheep using artificial neural networks. *Scientific Reports* **12**(1): 9009.

Aruna T S, Deb S M, Ruhil A P, Malhotra R and Yousaf S. 2022. Recent prediction tools for estimation of first lactation 305 days milk yield using monthly test days for genetic evaluation in crossbred cattle. *The Indian Journal of Animal Sciences* **92**(3): 374–377.

Balhara A K, Nayan V, Dey A, Singh K P, Dahiya S S and Singh I. 2017. Climate change and buffalo farming in major milk-producing states of India: Present status and need for addressing concerns. *Indian Journal of Animal Sciences* **87**(4): 403–411.

Balhara A K, Jan MH, Hooda E, Kumar K, Ghanghas A, Sangwan S, Balhara S, Phulia SK, Yadav S, Boora A and Kumar S. 2024. Prediction of core body temperature using infra-red thermography in buffaloes. *Italian Journal of Animal Science* **23**(1): 834–841.

Balhara S, Balhara A K, Dahiya N, Singh R P and Ruhil A P. 2022. An algorithm-based approach for identification of most relevant linear traits for selecting high-producing Murrah buffaloes. *Indian Journal of Animal Sciences* **92**(5): 618–623.

Dongre VB, Gandhi RS, Singh A and 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. *Livestock Science* **147**(1–3): 192–197.

Dahiya S, Kumar S and Kumar M. 2020. Current status of research on linear type traits in Indian cattle and future strategies. *Tropical Animal Health and Production* **52**: 2221–2232.

Fernández C, Soria E, Sanchez-Seiquer P, Gómez-Chova L, Magdalena R, Martín-Guerrero J D, Navarro M J and Serrano A J 2007. Weekly milk prediction on dairy goats using neural networks. *Neural Computing and Applications* **16**: 373–381.

Grzesiak W, Zaborski D, Pluciński M, Jędrzejczak-Silicka M, Pilarczyk R and Sablik P. 2025. The use of selected machine learning methods in dairy cattle farming: A review. *Animals* **15**(14): 2033.

Kaygisiz F and Sezgin F H. 2017. Forecasting goat milk production in turkey using artificial neural networks and Box-Jenkins models. *Animal Review* **4**(3): 45–52.

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): 1262–1268.

Lawal AI. 2020. An artificial neural network-based mathematical model for the prediction of blast-induced ground vibration in granite quarries in Ibadan, Oyo State, Nigeria. *Scientific African* **8**: e00413.

Murphy M D, O'Mahony M J, Shalloo L, French P and Upton J. 2014. Comparison of modelling techniques for milk-production forecasting. *Journal of Dairy Science* **97**(6): 3352–3363.

NDDB. 2017. Guidelines for type classification of cattle and buffalo. National Dairy Development Board, Anand, Gujarat. Available at: https://www.dairyknowledge.in/sites/default/files/animal_type_classification_guidelines_version_ii.pdf.

Python Software Foundation. 2021. Tkinter - Python Interface to Tcl/Tk (8.6.10) [Software]. Available at: https://docs.python.org/3/library/tkinter.html.Accessed on 15th Aug., 2023.

Rana E, Gupta A K, Singh A, Ruhil A P, Malhotra R, Yousuf S and Ete G. 2021. Prediction of first lactation 305-Day milk yield based on bimonthly test day milk yield records in Murrah buffaloes. *Indian Journal of Animal Research* **55**(4): 486–490.

Sahu S, Arya A, Kar D and Sarangi A. 2024. Correlation of Morphological Traits with Milk Production in Murrah Buffaloes. *Archives of Current Research International* **24**(5): 98–102.

Sangwan S, Jan MH, Hooda E, Choudhary R, Balhara S, Rautela R, Yadav S, Phulia SK, Sharma RK, Paul V and Pal Y, 2025. Urinary metabolite signatures as predictive biomarkers for estrus detection in water buffaloes: a proton-NMR based study. *Veterinary Quarterly* **45**(1): 2593361.

Sharma R, Dahiya S, Bangar YC, Hada R and Yadav DC. 2024. Genetic and phenotypic trends for production and reproduction traits in Murrah buffaloes. *Reproduction in Domestic Animals* **59**(9): 14712.

Singh N P, Usman S M, 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.

Slob N, Catal C and Kassahun A. 2021. Application of machine learning to improve dairy farm management: a systematic literature review. *Preventive Veterinary Medicine* **187**: 105237.

Usman S M, Singh N P, Dutt T, Tiwari R and Kumar A. 2020. Comparative study of artificial neural network algorithms' performance for prediction of FL305DMY in crossbred cattle. *Journal of Entomology and Zoology Studies* **8**: 516–520.

Waikato environment for knowledge analysis (WEKA). Available at: https://www.cs.waikato.ac.nz/ml/weka. Accessed on 18th Aug., 2023.