Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 67(1) 2013 1-11

Predicting Economic Traits in Murrah Buffaloes with Connectionist Models

A.K. Sharma, D.K. Jain, A.K. Chakravarty, R. Malhotra and A.P. Ruhil
National Dairy Research Institute, (*Deemed University*), *Karnal*

Received 3 May 2011; Revised 23 June 2011; Accepted 10 July 2011

SUMMARY

In this paper, several predictive models based on connectionist paradigms and conventional multiple regression approach are proposed to predict milk yield in different lactations as well as for overall data of Murrah buffaloes. The data pertaining to various economic traits including reproductive and productive characters are utilised for this purpose. The prediction potential of the connectionist models is compared with that of the conventional Multiple Linear Regression (MLR) models. The results revealed that the connectionist models developed in this study seem to be suitable as plausible alternative to conventional MLR models for predicting milk production in Murrah buffaloes.

Keywords: Connectionist model, Economic traits, Error back propagation, Generalised regression, Murrah buffalo, Prediction, Radial basis function.

1. INTRODUCTION

In Animal Improvement Programme, there is a need to develop the predictive model for prediction of the economic traits based on which the animals are selected or rejected for performing in the next generation in the herd. Conventional regression models have been used extensively for various prediction tasks in the field of dairying (Sundaresan *et al.* 1954, Puri and Sharma 1965, Schaeffer *et al.* 1977, Jain and Taneja 1984, Gandhi and Gurnani 1988, and Geetha *et al.* 2006). Generally, these models are based on certain assumptions, which are essential for their proper operation. However, there are some situations where assumptions get violated in practice.

Connectionist paradigm, comparatively a new branch of nonlinear techniques, is gaining momentum as potential alternative to conventional regression models for solving various real-life problems. However, there has been relatively little research into application of connectionist models in the field of agriculture in general and dairying in particular, especially in India. The research in this field is still at developmental stage across the globe. Majority of the studies showing application of connectionist models to different aspects of animal sciences and dairying have been conducted outside the country (*e.g.*, Salehi *et al.* 1998, Sanzogni and Kerr 2001, Kominakis *et al.* 2002, Cveticanin 2005, Grzesiak *et al.* 2006, Fernández *et al.* 2006, Fernández *et al.* 2007, Craninxa *et al.* 2008, Edriss *et al.* 2008, Cavero *et al.* 2009, Hassan *et al.* 2009, Hettinga *et al.* 2009, Grzesiak *et al.* 2010, Njubi *et al.* 2010, *etc.*). However, a few indigenous studies have recently been reported in the literature. These studies mainly focused on application of connectionist models for prediction of the first lactation 305-day milk yield in Karan-Fries dairy cattle (Sharma *et al.* 2006, Sharma *et al.* 2007, Sharma *et al.* 2010); to predict the lifetime milk production in Sahiwal Cattle (Gandhi *et al.* 2010) and for predicting the 305-day lactation milk yield of Sahiwal cows using partial lactation records of test days milk yield to rank the sires at younger age to improve the quality of milking animals and reducing generation intervals (Ruhil *et al.* 2011).

Corresponding author : A.K. Sharma
E-mail address : adeshkumar.sharma@gmail.com

It is evident from the foregoing review of literature that most of the research studies related to prediction of milk yield in dairy cattle are based on multivariate analysis such as multiple regression approach using various combinations of growth, reproduction and production traits. However, the attempt to compare the regression model *vis-à-vis* connectionist model is rather scanty and for the first time has been made in the world, for buffaloes. Therefore, the present study was undertaken to investigate the application potential of such an emerging soft computing paradigm as connectionist models in the prospective area of milk yield prediction in Murrah buffaloes for more effective decision- and policy-making for herd management.

1.1 Connectionist Models

The field of connectionism is a branch of cognitive science and has originated from diverse sources, ranging from the fascination of mankind with understanding and emulating the human brain, to broader issues of copying human abilities such as speech and the use of language, to the practical commercial, scientific, and engineering disciplines of pattern recognition, modelling, and prediction. Generally, connectionist models consist of layers of interconnected neurons, each neuron producing a nonlinear function of its input. The input to a neuron may come from other neurons or directly from the input data. Also, some neurons are identified with the output of the network. The complete network, therefore, represents a complex set of interdependencies, which may incorporate any degree of nonlinearity, allowing very general functions to be modelled. In the simplest connectionist networks, the output from one neuron is fed into another neuron in such a way so as to propagate the inherent features through layers of interconnecting neurons. It has been argued that connectionist models epitomise to a certain extent the behaviour of networks of neurons in the human brain. Connectionist modelling approaches combine the complexity of some of the statistical techniques with the machine learning objective of imitating human intelligence, however, this is done at a more 'unconscious' level; and hence, there is no accompanying ability to make learned concepts transparent to the user.

Feed-forward connectionist models allow signals to travel in one direction only, *i.e.*, from input towards output. These models can be considered as simple straightforward networks that associate inputs with outputs.

Connectionist models consist of the following three principal elements:

(a) Topology – the way a connectionist network is organised into layers and the manner in which these layers are interconnected;

(b) Learning – the technique by which information is stored in the network; and

(c) Recall – how the stored information is retrieved from the network.

The basic structure of a connectionist model consists of artificial neurons also sometimes referred to as processing elements (Fig. 1) and are analogous to biological neurons in the human brain, which are grouped into layers (or slabs). The most common connectionist structure consists of an input layer, one or more hidden layers and an output layer.

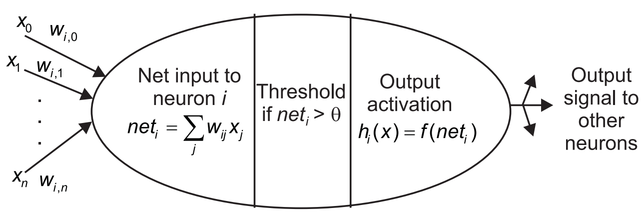


Fig. 1. Schematic representation of an artificial neuron.

Let the input dimension be n ($n \in Z_+$) and let the number of hidden neurons be m ($m \in Z_+$). Z_+ is the set of positive integers. The training pairs are represented by $D = \{x^{(p)}, t^{(p)}\}$, where $x^{(p)}$ and $t^{(p)}$ denote input and corresponding target patterns; $p = 1, 2, ..., P$; $P \in Z_+$, is the number of training exemplars; and the index p is always assumed to be present implicitly. The matrix **w** denotes the input to the hidden neurons connection strength, w_{ij} is the $(i, j)^{\text{th}}$ element of the matrix **w** representing the connection strength between the j^{th} input and the i^{th} hidden layer neuron. With this nomenclature, the net input to the i^{th} hidden layer neuron is given by

$$net_i = \sum_{j=1}^{n} w_{ij} x_j + \theta_i^{(1)} = \mathbf{w}_i \cdot \mathbf{x} + \theta_i^{(1)} \qquad (1.1)$$

where $\theta_i^{(1)}$ is the bias of the i^{th} hidden layer neuron. The output from the i^{th} hidden layer neuron is given by

$$h_i(\mathbf{x}) = f^{(1)}(net_i) \tag{1.2}$$

where $f^{(1)}(\cdot)$ is a nonlinear activation function.

The activation function/transfer function determines the output from a summation of the weighted inputs of a neuron. The activation functions for neurons in the hidden layer are often nonlinear and they provide the nonlinearities for the network. The choice of activation functions may strongly influence complexity and performance of connectionist models. Sigmoidal activation functions are most commonly used.

The net input to the output neuron may be defined similarly as eq. (1.1) as follows

$$net = \sum_{i=1}^{m} v_i h_i + \theta^{(2)} = \mathbf{v} \cdot \mathbf{h} + \theta^{(2)} \tag{1.3}$$

where v_i represents the connection strength between the i^{th} hidden layer neuron and the output neuron, while $\theta^{(2)}$ is the bias of the output neuron.

Adding a bias neuron x_0 with input value as +1, eq. (1.1) can be rewritten as

$$net_i = \sum_{j=0}^{n} w_{ij} x_j = \mathbf{W}_i \cdot \mathbf{x} \tag{1.4}$$

where $w_{i0} = W_{i0} = \theta_i^{(1)}$ and $\mathbf{W}_i$ is the weight vector $\mathbf{w}_i$ (associated with the i^{th} hidden neuron) augmented by the 0^{th} column corresponding to the bias. Similarly, introducing an auxiliary hidden neuron ($i = 0$) such that $h_0 = +1$, allows us to redefine eq. (1.3) as

$$net = \sum_{i=0}^{m} v_i h_i = \mathbf{V} \cdot \mathbf{h} \tag{1.5}$$

where $v_0 \equiv \theta^{(2)}$.

The equation for the network output neuron is given by

$$net_o = f^{(2)}(net) = net \tag{1.6}$$

where $f^{(2)}(\cdot)$ is a linear function.

The notations are diagrammatically exemplified in Fig. 2. This figure represents an n-input, m-hidden neuron and one-output feed-forward connectionist network. Such a connectionist model is trained to fit a dataset D by minimising an error function (or performance function) as

$$F = E_D(\mathbf{W}) = \frac{1}{P}\sum_{p=1}^{P} \varepsilon^2 = \frac{1}{P}\sum_{p=1}^{P}\left(net_o^{(p)} - t^{(p)}\right)^2 \tag{1.7}$$

This function is minimised using any standard optimisation method.

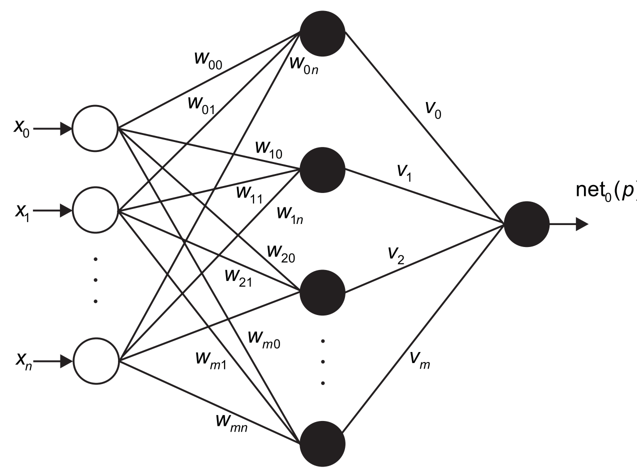


Fig. 2. Schematic of a feed-forward connectionist model.

1.1.1. Error Back Propagation Connectionist Model

The Error Back Propagation (EBP) algorithm *ab initio* is a gradient descent algorithm in which the network weights are moved along the negative of gradient of performance function. The term error back propagation refers to the manner in which the gradient is computed for nonlinear multilayer networks. There are a number of variations of basic algorithm that are based on other heuristics and standard optimisation techniques, such as variable learning rate gradient descent, conjugate gradient and Newton methods.

The simplest implementation of EBP algorithm updates the network weights and biases in the direction in which the performance function decreases most rapidly, *i.e.*, the negative of gradient. The $(k+1)^{th}$ iteration of this algorithm (Demuth and Beale 2004) is given by:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{g}_k \tag{1.8}$$

where $\mathbf{x}_k$ is a vector of current weights and biases, α_k is the learning rate, and $\mathbf{g}_k$ is the current gradient. The EBP method used in this study is gradient descent algorithm with variable learning rate (Vogl *et al.* 1988), which is based on a heuristic technique as delineated in the next sub-section.

1.1.1.1 Variable learning rate gradient descent algorithm

The rules governing the algorithm are:

(a) If the squared error (over the entire training set) increases by more than some set percentage ζ (typically one to five percent) after a weight update using eq. (1.8), then the weight update is discarded and the learning rate is multiplied by some factor ρ $(0 \leq \rho < 1)$.

(b) If the squared error decreases after a weight update, then the weight update is accepted and the learning rate is multiplied by some factor $\eta > 1$.

(c) If the squared error increases by less than ζ, then the weight update is accepted but learning rate is unchanged.

1.1.2 Radial Basis Function Connectionist Models

A Radial Basis Function (RBF) connectionist model is a standard three-layer feed-forward connectionist model that comprises of an input layer, a hidden layer of RBF neurons and output layer of linear neurons. The input layer transfers the input vector **x** to the hidden neurons. Generally, there are as many hidden neurons as the number of input and target samples in the training set, *i.e.,* one hidden neuron is assigned to each input vector, $\mathbf{x}^{(p)} = \left(x_1^{(p)}, x_2^{(p)}, ..., x_n^{(p)}\right)$, p = 1, 2, ..., P. This form of RBF connectionist model is known as regularisation network.

As such, there will be P hidden neurons and each input vector generates a centre vector for one hidden neuron. Such a regularisation network becomes very large as the number of observations or input vector increases, and thus leading to a computation intensive process. A clustering method can be applied to the input vectors to reduce the size of the network. There exists several clustering methods, *e.g.*, fixed centres; flexible centres. The one used in this study is a variant of the fixed centres method and is briefly described below.

In this method, we specify a maximum number of hidden neurons, a target error level and a spread width. The model starts with zero hidden neurons and adds one hidden neuron and one corresponding centre vector at a time until either the error falls beneath the target level, or the maximum number of hidden neurons has been reached (Demuth and Beale 2004). At each step the model chooses the particular input vector as its centre that lowers the network error the most. The value of spread constant σ is determined through trial and error in the present study.

The transformation process in Fig. 2 takes the following form. Each hidden layer neuron, φ_i generates a centre vector, $\mathbf{c}_i = \left(c_1^{(i)}, c_2^{(i)}, ..., c_n^{(i)}\right)$, and calculates the distance between the input vector (or clusters) and the centre vector. As the distance measure used in this study is Euclidean norm, therefore, the result for the hidden neuron i, net_i is given as

$$net_i = \|\mathbf{x} - \mathbf{c}_i\| = \left[\sum_{j=1}^{n}\left(x_j - c_{ij}\right)^2\right]^{1/2}, \quad i = 1, 2, \;..., \; m \qquad (1.9)$$

where $m \leq P$ is the maximum number of hidden layer neurons.

The vector **net** is then transformed in the output neuron by an activation function. The activation function can be taken from a variety of functions on the non-negative real numbers that takes its maximum value of zero for a net_i, and approaches zero as net_i approaches infinity. The Gaussian basis function is used as activation function in the present study as the computations are not that complex if a Gaussian kernel is employed. The effective range of kernels is determined by the values allocated to the centre and spread width of radial basis function. The Gaussian function used is given as

$$\varphi(net_i) = \varphi(\|\mathbf{x} - \mathbf{c}_i\|) = e^{-\left(net_i^2/2\sigma_i^2\right)} \qquad (1.10)$$

where σ_i is the size of the spread width (or receptive field) of the neuron i, *i.e.*, the range of the values in which the output has a significant response to the input **x**. The radial basis function φ is radially symmetrical around the centre, *i.e.*, it takes the maximum value as unity when the input vector **x** is identical to the centre vector **c**, and decreases dramatically when **x** has diverged from the centre by as much as the spread width. The spread width is similar to standard deviation, and the whole activation function is analogous to a normal probability distribution function with mean **c** and standard deviation σ_i.

The final output generated by the RBF connectionist model is a linear function of the hidden layer outputs and is represented as

$$net = \sum_{i=1}^{m} v_i \varphi\left(\left\|\mathbf{x} - \mathbf{c}_i\right\|\right) + \theta_i^{(2)} \quad (1.11)$$

The parameters of the model to be estimated are the matrix **V** of weights and biases for output layer neurons; and vectors **σ** and **c**. The values of **V** are determined using a learning rule whereas the values of **σ** and **c** are determined by the clustering method as delineated above within this section.

For every training example, we can compute the actual network output net_o and error $\varepsilon^{(p)}$, $p = 1, 2, ..., P$. The goal of RBF network learning is to minimise the error function defined by eq. (1.7).

1.1.3. Generalised Regression Connectionist Model

A nonlinear regression model can be given by

$$t^{(p)} = f(\mathbf{x}^{(p)}) + \in^{(p)}, \; p = 1, 2, ..., P \quad (1.12)$$

where $f(\mathbf{x})$ is the unknown regression function. We know that $f(\mathbf{x})$ can also be written as the conditional mean of the t given $\mathbf{x}$ (*i.e.*, the regression of t on $\mathbf{x}$) as below (Haykin 2001).

$$f(\mathbf{x}) = E(t \mid \mathbf{x}) \quad (1.13)$$

The definition of the expectation of a random variable gives

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} t \, f_T(t \mid \mathbf{x}) dt \quad (1.14)$$

where f_T is the conditional Probability Density Function (PDF) of t given $\mathbf{x}$. We also know that

$$f_T(t \mid \mathbf{x}) = \frac{f_{\mathbf{X},T}(\mathbf{x}, t)}{f_{\mathbf{X}}(\mathbf{x})} \quad (1.15)$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the PDF of $\mathbf{x}$ and $f_{\mathbf{X},T}(t \mid \mathbf{x})$ is the joint PDF of $\mathbf{x}$ and t. Using these, we can get

$$f(\mathbf{x}) = \frac{\int_{-\infty}^{\infty} t \, f_{\mathbf{X},T}(\mathbf{x}, t) dt}{f_{\mathbf{X}}(\mathbf{x})} \quad (1.16)$$

Here, the probability density functions $f_{\mathbf{X},T}$ and $f_{\mathbf{X}}$ are not known and these are to be estimated from the training dataset. Now, following Haykin (2001) and using the theory of kernel functions associated with PDF, we can write

$$\hat{f}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{P \cdot \sigma^n} \sum_{p=1}^{P} K\left(\frac{\mathbf{x} - \mathbf{x}^{(p)}}{\sigma}\right) \text{ for } \mathbf{x} \in \Re^n \quad (1.17)$$

and

$$\int_{-\infty}^{\infty} t \, \hat{f}_{\mathbf{X},T}(\mathbf{x}, t) \, dt = \frac{1}{P \cdot \sigma^n} \sum_{p=1}^{P} t^{(p)} K\left(\frac{\mathbf{x} - \mathbf{x}^{(p)}}{\sigma}\right) \quad (1.18)$$

Thus,

$$\hat{f}(\mathbf{x}) = \Gamma(\mathbf{x}) = \frac{\sum_{p=1}^{P} t^{(p)} K\left(\frac{\mathbf{x} - \mathbf{x}^{(p)}}{\sigma}\right)}{\sum_{p=1}^{P} K\left(\frac{\mathbf{x} - \mathbf{x}^{(p)}}{\sigma}\right)} \quad (1.19)$$

In the present study, we have used $K(\mathbf{x})$ as the multivariate Gaussian distribution function given by

$$K(\mathbf{x}) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \cdot e^{-\left(\frac{\|\mathbf{x}\|^2}{2}\right)} \quad (1.20)$$

As such, we can write that

$$\Gamma(\mathbf{x}) = \frac{\sum_{p=1}^{P} t^{(p)} \cdot e^{-\left(\frac{\left\|\mathbf{x} - \mathbf{x}^{(p)}\right\|^2}{2\sigma^2}\right)}}{\sum_{p=1}^{P} e^{-\left(\frac{\left\|\mathbf{x} - \mathbf{x}^{(p)}\right\|^2}{2\sigma^2}\right)}} \quad (1.21)$$

The resulting regression, eq. (1.21), also known as Nadaraya-Watson kernel regression estimator, is directly applicable to problems involving numerical data. The estimate $\Gamma(\mathbf{x})$ can be visualised as an exponentially weighted average of all the observed values, $t^{(p)}$, where each observed value is weighted exponentially according to its Euclidean distance from $\mathbf{x}$.

Connectionist model implementation

General Regression (GR) connectionist model implementation was first introduced by Specht (1991). Generalised regression connectionist model is the paradigm of a RBF connectionist model, which is a feed-forward network that uses single hidden layer, generally with a Gaussian response function that is radially symmetric. Let w_{ij} be the desired output of the input training vector $\mathbf{x}^{(j)}$ and i^{th} output. Now eq. (1.21) can be expressed as follows

$$net_i = \frac{\sum_{j=1}^{P} \varphi_j w_{ij}}{\sum_{j=1}^{P} \varphi_j} \tag{1.22}$$

where

$$\varphi_j = e^{-\left(\frac{d_i^2}{2\sigma^2}\right)} \tag{1.23}$$

and

$$d_i^2 = (\mathbf{x} - \mathbf{x}^{(i)})^T (\mathbf{x} - \mathbf{x}^{(i)}).$$

According to eq. (1.22), the topology of a GR connectionist model (Fig. 3) can be described as per the following points.

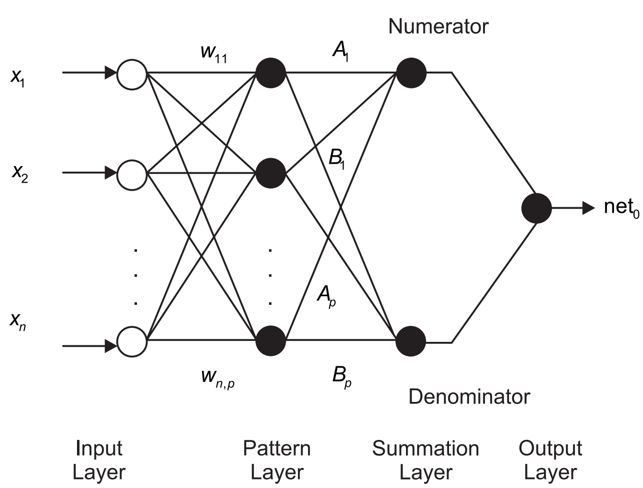


Fig. 3. Basic architecture of a generalised regression connectionist model.

(i) The input layer (input cells), which is fully connected to the pattern layer;

(ii) The pattern layer which has as many neurons as there are input and target pairs in training set. Specifically, the first layer weights are set to $\mathbf{x}'$. It computes the pattern functions φ_j, which is obtained from eq. (1.23);

(iii) The summation layer, which has two neurons, *viz.*, numerator (A) and denominator (B). The first neuron has input weights equal to number of input and target pairs, but in this case the weights are set to the target vector, t. Now, it computes the numerator by summation of the exponential terms multiplied by the $t^{(p)}$ associated with $\mathbf{x}^{(p)}$. The second neuron has input weights equal to 1, then the denominator is the summation of exponential terms alone; and

(iv) Finally, the output neuron divides the numerator by the denominator to provide the prediction result net_o.

Variable σ can be adjusted to provide different levels of function smoothing. Larger values for σ cause the estimated function to be smoother than a function estimated using lower values of σ. Choosing the right value for the smoothing parameter requires a certain amount of experimentation.

2. MATERIALS AND METHODS

2.1 Data

The data pertaining to various economic traits on Murrah buffaloes for the period 1990-2006, being maintained at Institute Livestock Farm, NDRI, Karnal (India), including pedigree information; economic traits: reproductive characters – age at maturity, age at calving (up to first six calvings), service period, gestation period, calving interval, number of services per conception and individual conception rate; and production traits – lactation 305-day milk yield (1-5 lactations), lactation length and dry period were used.

2.2 Data Preprocessing

The data were subjected to Least-Squares analysis. Further, the comparison among sub-classes within years, seasons in each lactation and also parities for pooled lactation was made. The constants for significant effects of non-genetic factors on reproduction and production traits were used for adjusting the data. Furthermore, regression analysis technique was employed on the final adjusted dataset with all possible combinations of the different economic traits (lactation wise as well as overall) so as to isolate representative traits affecting the milk production of Murrah buffaloes. Based on the higher associations of predictor traits with the predicted trait, the predictor traits have been included in various models as shown in Table 1.

2.2.1 Design and Development of Connectionist and Regression Models

Three feed-forward connectionist models based on three learning paradigms, *viz.*, EBP, RBF and GR learning algorithms have been proposed to predict 305-day milk yield for each of the different lactations (up to first six calvings) as well as for the overall data on Murrah buffaloes using the datasets shown in Table 1. Hence, there are as many as eighteen models developed in this study.

Various combinations of several architectural parameters such as number of hidden layers, number of neurons in each hidden layer, transfer function for each hidden layer, learning rate, error goal, epochs, spread constant, regularisation constant, *etc*., were empirically explored to reach an optimum configuration for each connectionist model. Linear function was used as transfer function for the output layer for all the connectionist models. The Neural Network Toolbox under MATLAB software was used to carry out all the training and simulation experiments. Further experimental details have been described for each model (lactation wise) in the following Section 3.

Table 1. Traits included in the datasets for development of Connectionist models *vis-à-vis* Conventional Regression models for predicting 305-day milk yield in different lactations as well as for overall data of Murrah buffaloes.

Dependent Trait	Independent Traits								
First Lactation 305-day Milk Yield (FLMY305) (N=415)	Age at Maturity (AAM)	First Breeding Interval (FBI)							
Second Lactation 305-day Milk Yield (SLMY305) (N=268)	AAM	FLMY 305	First Dry Period (FDP)	First Lactation Length (FLL)					
Third Lactation 305-day Milk Yield (TLMY305) (N=161)	FLMY 305	FDP	FLL	Second Service Period (SSP)	SLMY 305	Second Dry Period (SDP)	Second Lactation Length (SLL)	Third Breeding Interval (TBI)	
Fourth Lactation 305-day Milk Yield (FoLMY305) (N=99)	FLMY 305	SLMY 305	SLL	Third Service Period (TSP)	TLMY 305	Third Dry Period (TDP)	Third Location Length (TLL)		
Fifth Lactation 305-day Milk Yield (FiLMY305) (N=54)	SLMY 305	TSP	TLMY 305	TDP	TLL	Fourth Service Period (FSP)	FoLMY 305	Fourth Dry Period (FoDP)	Fourth Location Length (FoLL)
Pooled Lactation 305-day Milk Yield (PLMY305) (N=599)	AAM	Breeding Interval (BI)	Service Period (SP)	Dry Period (DP)					

3. RESULTS AND DISCUSSION

3.1 Predicting First Lactation 305-Day Milk Yield

Many models were developed using various combinations of independent traits for predicting dependent traits, *i.e.*, 305-day milk production. Also, a Multiple Linear Regression (MLR) model was fitted corresponding to each set of the three connectionist models using same dataset and data partitioning scheme for comparing performance of the connectionist models. The results have been summarised and presented in Table 2.

Table 2. Performance of Connectionist models *vis-à-vis* Conventional Regression models for predicting 305-day milk yield in different lactations as well as for pooled data of Murrah Buffaloes.

Lactation	RMSE (%)			
	Multiple Linear Regression Model	Connectionist Models		
		EBP	RBF	GR
First Lactation	52.49	**51.49**	57.64	85.71
Second Lactation	38.94	42.78	**36.25**	41.36
Third Lactation	27.13	23.37	23.89	**21.90**
Fourth Lactation	34.22	32.16	**31.88**	36.79
Fifth Lactation	31.32	30.87	23.78	**19.28**
Pooled/Overall Lactation	22.12	**20.87**	21.61	22.97

Three connectionist models using EBP, RBF and GR learning algorithms were developed to predict FLMY305 using various combinations of architectural parameters. The best model (shown as bold typeface in Table 2) was empirically established to be founded on EBP through gradient descent method with adaptive learning rate comprising of two hidden layers, each containing two neurons and having log-sigmoid function employed as the transfer function; epochs, error goal and learning rate being arbitrarily set to 5000, 0.01 and 0.01, respectively. The dataset was split randomly into two disjoint subsets, *viz.*, 'training set' containing 80 per cent of the patterns (used for training the connectionist model) and 'test set' containing remaining 20 per cent data points (used for validating the connectionist model). The Root Mean Square Error (RMSE) for the best model was 51.49 per cent (Fig. 4a). Also, an MLR model was fitted using the same training set, which was validated on the same test set. The resultant RMSE was as 52.49 per cent (Fig. 4a).

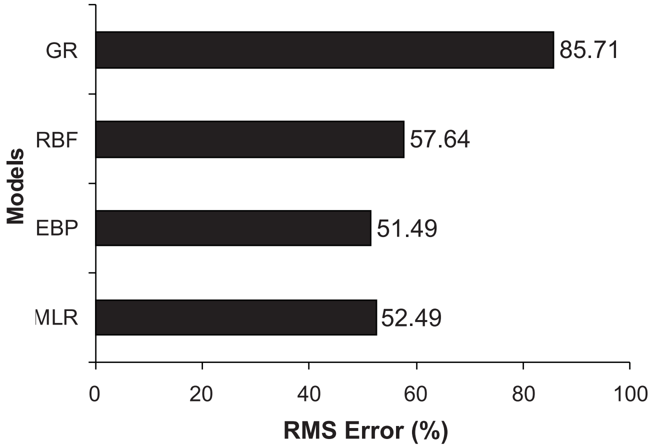
3.2 Predicting Second Lactation 305-Day Milk Yield

Three connectionist models using EBP, RBF and GR learning algorithms were developed to predict SLMY305 using various combinations of several architectural parameters. The best model (shown as bold typeface in Table 2) was empirically found to be based on RBF learning algorithm containing four neurons in the hidden layer; epochs, error goal and spread constant being set to 4, 0.001 and 215, respectively. The dataset was split randomly into two disjoint subsets, *viz.*, 'training set' containing 90 per cent of the patterns (used for training the connectionist model) and 'test set' containing remaining 10 per cent data points (used for validating the connectionist model). The RMSE for the best model was 36.25 per cent (Fig. 4b). Also, an MLR model was fitted using the same training set, which was validated on the same test set. The resultant RMSE was as 38.94 per cent (Fig. 4b).

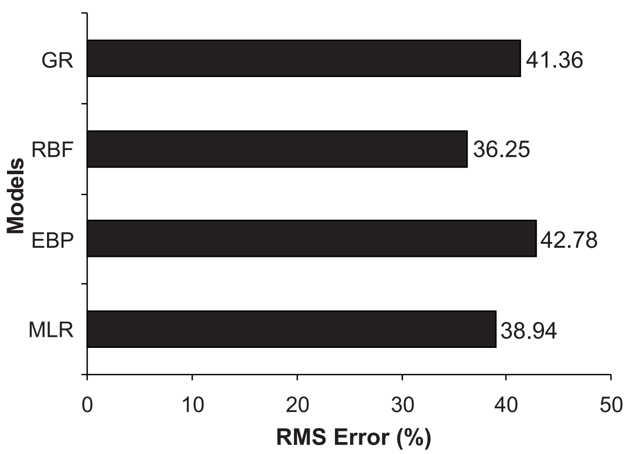
3.3 Predicting Third Lactation 305-Day Milk Yield

Three connectionist models using EBP, RBF and GR learning algorithms were developed to predict TLMY305 using various combinations of several architectural parameters. The best model (shown as bold typeface in Table 2) was empirically found to be based on GR learning algorithm containing 129 neurons in the hidden layer; epochs and spread constant being set to 129 and 500, respectively. The dataset was split randomly into two disjoint subsets, *viz.*, 'training set' containing 80 per cent of the patterns (used for training the connectionist model) and 'test set' containing remaining 20 per cent data points (used for validating the connectionist model).

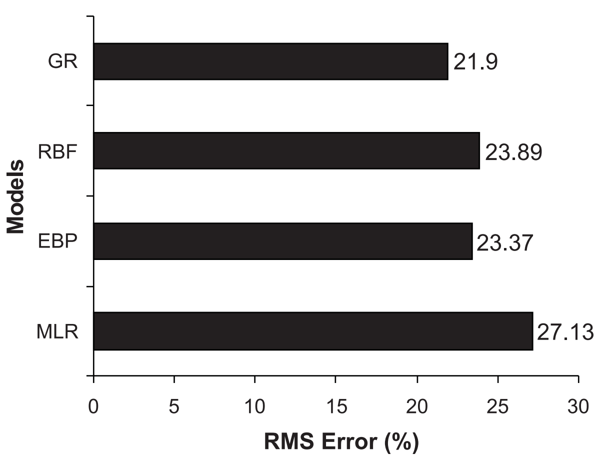
The RMSE for the best model was 21.90 per cent (Fig. 4c). Also, an MLR model was developed using the same training set, which was validated on the same test set. The resultant RMSE was as 27.13 per cent (Fig. 4c).



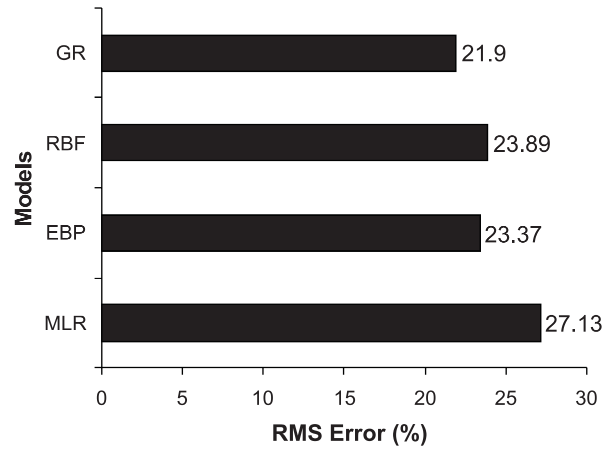
(a) First lactation milk yield prediction



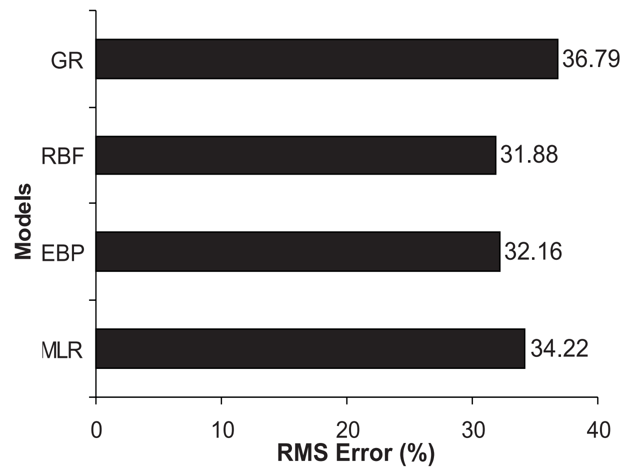
(b) Second lactation milk yield prediction



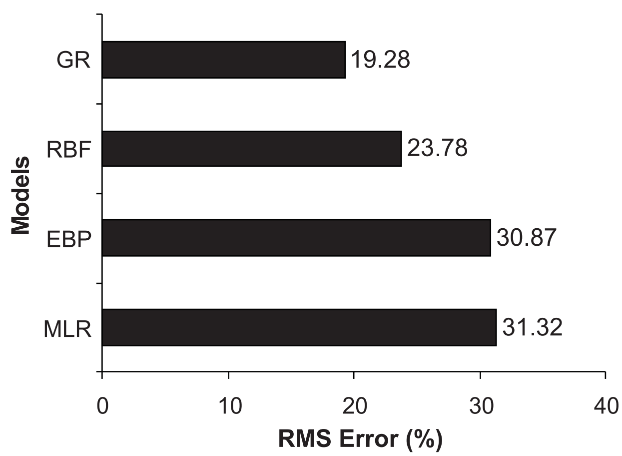
(c) Third lactation milk yield prediction



(d) Fourth lactation milk yield prediction



(e) Fifth lactation milk yield prediction



(f) Pooled lactation milk yield prediction

Fig. 4. Performance comparison of the connectionist and conventional regression models to predict 305-day milk yield in different lactations (up to six calvings) and in pooled lactation of Murrah buffaloes.

3.4 Predicting Fourth Lactation 305-Day Milk Yield

Three connectionist models using EBP, RBF and GR learning algorithms were developed to predict FoLMY305 using various combinations of several architectural parameters. The best model (shown as bold typeface in Table 2) was empirically found to be based on RBF learning algorithm containing two neurons in the hidden layer; epochs, learning rate and spread constant being set to two, 0.0001 and 2200, respectively. The dataset was split randomly into two disjoint subsets, *viz.*, 'training set' containing 80 per cent of the patterns (used for training the connectionist model) and 'test set' containing remaining 20 per cent data points (used for validating the connectionist model). The RMSE for the best model was 31.88 per cent (Fig. 4d). Also, an MLR model was developed using the same training set, which was validated on the same test set. The resultant RMSE was as 34.22 per cent (Fig. 4d).

3.5 Predicting Fifth Lactation 305-Day Milk Yield

The same set of three connectionist models using EBP, RBF and GR learning algorithms were developed to predict FiLMY305 using various combinations of several architectural parameters. The best model (shown as bold typeface in Table 2) was empirically found to be based on GR learning algorithm containing 44 neurons in the hidden layer; epochs and spread constant being set to 44 and 250, respectively. The dataset was split randomly into two disjoint subsets, *viz.*, 'training set' containing 80 per cent of the patterns (used for training the connectionist model) and 'test set' containing remaining 20 per cent data points (used for validating the connectionist model). The RMSE for the best model was 19.28 per cent (Fig. 4e). Also, an MLR model was fitted using the same training set, which was validated on the same test set. The resultant RMSE was as 31.32 per cent (Fig. 4e).

3.6 Predicting Pooled Lactation 305-Day Milk Yield

Three connectionist models using learning algorithms as stated earlier were developed to predict PLMY305 using various combinations of several architectural parameters. The best model (shown as bold typeface in Table 2) was empirically established to be founded on EBP through Levenberg–Marquardt method comprising of a single hidden layer containing ten neurons and having log-sigmoid function employed as the transfer function; epochs, error goal and learning rate being arbitrarily set to 5000, 0.001 and 0.01, respectively. The dataset was split randomly into two disjoint subsets, *viz.*, 'training set' containing 80 per cent of the patterns (used for training the connectionist model) and 'test set' containing remaining 20 per cent data points (used for validating the connectionist model). The RMSE for the best model was 20.87 per cent (Fig. 4f). Also, an MLR model was fitted using the same training set, which was validated on the same test set. The resultant RMSE was as 22.12 per cent (Fig. 4f).

4. CONCLUSION

In this paper, eighteen predictive models based on connectionist paradigms, *viz.*, error back propagation, radial basis function and generalised regression learning algorithms have been developed to predict 305-day milk yield in different lactations (up to first six calvings) as well as for overall data of Murrah buffaloes. Various combinations of several architectural parameters such as number of hidden layers, number of neurons in each hidden layer, transfer function for each hidden layer, learning rate, error goal, epochs, spread constant, regularisation constant, *etc*., were empirically explored so as to reach an optimum configuration for each connectionist model. The results of this study revealed that the connectionist models have relatively better potential over the conventional MLR technique. Hence, it is concluded that the connectionist models described in this paper are found to be suitable as plausible alternative to conventional MLR models for predicting 305-day milk production in Murrah buffaloes.

REFERENCES

Cavero, D., Tölle, K.H., Henze, C., Buxadé, C. and Krieter, J. (2008). Mastitis detection in dairy cows by application of neural networks. *Livest. Sci.*, **114**, 280-286.

Craninxa, M., Fieveza, V., Vlaemincka, B. and De Baets, B. (2008). Artificial neural network models of the rumen fermentation pattern in dairy cattle. *Comput. Electron. Agr.*, **60**, 226-238.

Cveticanin, D. (2005). *Modelling and Simulation of Cow Locomotion for Dynamic Weighing in Modern Dairy Farming* (PhD Thesis). Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt, Technische Universität München.

Demuth, H.B. and Beale, M. (2004). *User's Guide for Neural Network Toolbox (version 4) for use with MATLAB 6.1*. The MathWorks Inc., Natick, MA.

Edriss, M.A., Hosseinnia, P., Edrisi, M., Rahmani, H.R. and Nilforooshan, M.A. (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.

Fernández, C., Soria, E., Martín, J.D. and Serrano, A.J. (2006). Neural networks for animal science applications: two case studies. *Expert Syst. Appl.*, **31**, 444-450.

Fernández, C., Soria, E., Sánchez-Seiquer, P., Gómez-Chova, L., Magdalena, R., Martín, J.D., Navarro, M.J. and Serrano, A.J. (2007). Weekly milk prediction on dairy goats using neural networks. *Neural Comput. Appl.*, **17**, 373-381.

Gandhi, R.S. and Gurnani, M. (1988). Optimum combinations of early traits for prediction of 8-year lifetime production in Sahiwal cattle. *Asian J. Dairy Res.*, **7**, 7-12.

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. *J. Appl. Anim. Res.*, **38**, 233-237.

Geetha, E., Chakravarty, A.K. and Vinaya-Kumar, K. (2006). Genetic persistency of first lactation milk yield estimated using random regression model for Indian Murrah Buffaloes. *Asian Austral. J. Anim.*, **19**, 1696-1701.

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 (ANNs). *Comput. Electron. Agr.*, **54**, 69-83.

Grzesiak, W., Zaborski, D., Sablik, P., Zukiewicz, A., Dybus, A. and Szatkowska, I. (2010). Detection of cows with insemination problems using selected classification models. *Comput. Electron. Agr.*, **74**, 265-273.

Hassan, K.J., Samarasinghe, S. and Lopez-Benavides, M.G. (2009). Use of neural networks to detect minor and major pathogens that cause bovine mastitis. *J. Dairy Sci.*, **92**, 1493-1499.

Haykin, S. (2001). *Neural Networks – A Comprehensive Foundation*. 2nd Edition. Pearson Education (Singapore) Pte. Ltd., Indian Branch, Delhi.

Hettinga, K.A., van Valenberg, H.J.F., Lam, T.J.G.M. and van Hooijdonk, A.C.M. (2009). The origin of the volatile metabolites found in mastitis milk. *Vet. Microbiol.*, **137**, 384-387.

Jain, A. and Taneja, V.K. (1984). Prediction of lifetime production in Murrah buffaloes. *Indian J. Anim. Sci.*, 70, 304-306.

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. *Comput. Electron. Agr.*, **35**, 35-48.

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. *Trop. Anim. Health Pro.*, **42**, 639-644.

Puri, T.R. and Sharma, K.N.S. (1965). Prediction of lifetime production on the basis of first lactation yield and age at first calving for selection of dairy cattle. *J. Dairy Sci.*, **48**, 462-467.

Ruhil, A.P., Gandhi, R.S., Monalisa, D., Behra, K. and Raja, T.V. (2011). Prediction of lactation yield based on partial lactation records using artificial neural networks. *Proc. 5th National Conference on Computing for Nation Development, INDIACom-2011*, Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi, India.

Salehi, F., Lacroix, R. and Wade, K.M. (1998). Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Comput. Electron. Agr.*, **20**, 199-213.

Sanzogni, L. and Kerr, D. (2001). Milk production estimates using feed forward artificial neural networks. *Comput. Electron. Agr.*, **32**, 21-30.

Schaeffer, L.R., Minder, C.E., McMillan, I. and Burnside, E.B. (1977). Non-linear techniques for predicting 305-day lactation production of Holsteins and Jerseys. *J. Dairy Sci.*, **60**, 1636-1644.

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 Comput. Appl.*, **15**, 359-365.

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. *Appl. Soft Comput.*, **7**, 1112-1120.

Sharma, P.C., Sharma, A.K., Gupta, A.K., Mohanty, T.K., Chakravarty, A.K. and Singh, A. (2010). Predicting lifetime performance in Karan-Fries dairy cattle using first lactation performance and functional traits – a connectionist approach. *Int. J. Comput. Int. Res.*, **6**, 557-560.

Specht, D.F. (1991). A general regression neural network. *IEEE T. Neural Networ.*, **2,** 568-576.

Sundaresan, D., Eldridge, F.E. and Atkeson, F.W. (1954). Age at first calving used with milk yield during first lactation to predict lifetime production of Indian cattle. *J. Dairy Sci.*, **37**, 1273-1282.

Vogl, T.P., Mangis, J.K., Zigler, A.K., Zink, W.T. and Alkon, D.L. (1988). Accelerating the convergence of the back-propagation method. *Biol. Cybern.*, **59**, 256-264.