African Journal of Biotechnology Vol. 9(30), pp. 4822-4825, 26 July, 2010
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 ©2010 Academic Journals

Full Length Research Paper

Predictions of semen production in ram using phenotypic traits by artificial neural network

Ali Ahmad Alaw Qotbi[1]*, Poorya Hossein Nia[2], Alireza Seidavi[1] and Shahrokh Ghovvati[2]

[1]Department of Animal Science, Islamic Azad University, Rasht Branch, Rasht, Iran.
[2]Department of Animal Science, Mashhad University of Ferdowsi, Mashhad, Iran

Accepted 21 June, 2010

Concentration of semen production is the most important fertility trait in ram and dimension of testis is a good criterion for identifying the quantity of semen production. Thus, prediction of that trait has important beneficial effect on the timely identification of genetically superior animals. Artificial neural network (ANN) system can be used as a decision making support system in ram industry as well as other industries. It can help breeders to predict future semen production based on phenotypic trait. Data from 24 rams of zandi breed in Tehran, Iran, were used. From 192 available data of phenotypic and semen concentration, 184 records were used for training a back propagation ANN system and 8 randomly chosen record (not used in the training process) were introduced to the trained neural network for evaluation. The result of the simulation showed that there was no significant difference between the observed and the predicted semen production ($p > 0.05$). The major use of this predictive system is to make accurate selection decision which is based on prior knowledge of the outcomes.

Key words: Artificial neural network, correlation, semen production, ram.

INTRODUCTION

Semen production (SP) in rams involves complicated and linear as well as non-linear interaction between genetic and environmental effect (Komonakis et al., 2002), for example, the number of spermatozoids production in relation with testis volume. In other words, rams present the season oscillation in their sexual behavior, hormonal action, testis volume and testis weight (Blache et al., 2000). Concentration of semen production is the most important fertility trait in ram and dimension of testis is a good criterion for identifying the quantity of semen production. Also, testis dimension in males has positive genetic relation with ovulation in ewes. So prediction of that trait has important economic, management and breeding points of view and makes use of prospective high strong rams which improve the farmers' economic proficiency. Also, much of the selection of superior rams is based on the ability of SP rams (Hosseinnia et al., 2007). Therefore, the sooner these rams can be identified, the sooner the collection of semen and marketing can be processed (Salehi et al., 1998, b). Accuracy rate of finding high ability rams is important. This is because feeding, breeding, maintenance, veterinary and other cost can be saved from superior and mis-caulling rams of high genetic value but a good source of gene pool will be lost.

The ability of artificial neural network (ANN) to detect patterns that relate input variables to their corresponding outputs in complex biological systems has resulted in some impressive success in classification and prediction (Wasserman et al., 1993). This has led to an increase of ANN application in different fields of animal science (Lacroix et al., 1995; Salehi et al., 1998b; Kominakis et al., 2002; Hosseinnia et al., 2007). Actually, ANN is a form of simulated human central nervosa system (Adamczyk et al., 2005) which is the same as biological neural network and is made up of sets of neurons. These neurons process the presented input and matching output in supervised manner and make extract non-linear relationship between input and output. Information processing

*Corresponding author. E-mail: aqotbi@iaurasht.ac.ir.

Abbreviations: SP, Semen production; **ANN,** artificial neural network; **RMSE,** root mean square error.

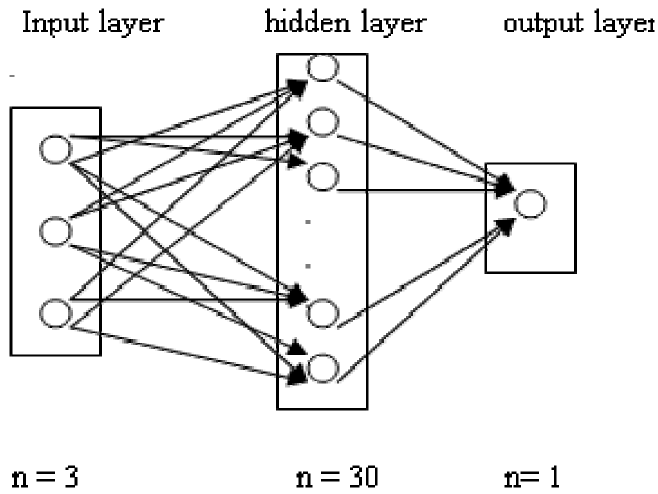

Figure 1. Neuron connection in an ANN system.

in ANN has a parallel form. Although mathematical or statistical models have wide ranges of application in agriculture, they have some inherent restriction. Identifying patterns and extract relationship between input and corresponding output in sample data through the learning process depends on the facts that the optimal net performance is dependent on recognizing and extracting non-linear relationship (Lacroix et al., 1995). Using this relation in the simulation stage, ANN can anticipate the output of the problem in complex biological system from known input. The pattern recognition ability of network may be improved by different technique. Common methods of improving network performance include finding an optimum network architecture and appropriate number of training cycles, using different input combinations (Yang et al., 1999) and using the learning parameter values (Salehi et al., 1998a). In addition, these methods can be useful in using input and output variable with high biological relationship (Hosseinnia et al., 2007). Also selection and pre-processing of data have been shown to affect network performance significantly (Ruan et al., 1997). However divergence in performance was often reported in array agriculture, due to over modification of net structure (Sabalani et al., 1995; Yang et al., 1999). It can be useful in using input and output variable with high biological relationship (Hosseinnia et al., 2007). Proportion of data category in training period has effect on the network performance (Lacroix et al., 1997), thus for appropriating optimal learning, a good data presentation is important.

The aim of this study is to predict rams' ability in semen production in two volume criteria using valuable phenoltypic information and investigating the ability and accuracy of ANN assessment in predicting rams' semen production ability for the selection of superior rams as prospective producers and the parent of the next generation. Also, the study investigates the proportion of any parameter from input variable on network response.

MATERIALS AND METHODS

Data was made available by the Animal Husbandry Division, Tehran University, Tehran, Iran. Collected SP trait record from 24 zandi rams were selected randomly for this study; semen was collected by artificial vaginal technique in summer and fall season. The sample data was considered, 184 records from testis volume, testis dimension and testis circumference were used as a phenoltypic variable and SP concentration, and were categorized into two set of semen concentration of 100 mm (sc) and total semen concentration (tsc). In order to train ANN for any of the SP category made of ANN (ANN1 and ANN2 respectively), three phenotypic variable and SP variable corresponding to individual ram were introduced to the system as input and output variable, respectively. Then the minimum and maximum values of each variable were mapped to the mean and standard deviation of 0 and 1, respectively.

Out of 192 records, 184 records were used for ANN training and 8 records were selected randomly for testing the simulated system. The whole data used for training were divided into 3 sub categories: 50% were used to record training set, 25% were used to record evaluation set and 25% of records as a testing set. Then these data were introduced to ANN as matrix in which any column excites a variable.

In order to construct the network, the neural network toolbox was used (MATLAB, 2006). The constructed network was a back propagation ANN with three layers of hidden input and output. The layers had 3, 30 and 1 neurons, respectively. The number of neuron in input and output layer must be exactly an equal number of variables in those layers. Figure 1 shows how neurons in different layers of ANN connect together. The tangent sigmoid transfer function was applied for input and hidden layers and the pure line transfer function was applied for output layer (MATLAB, 2006). The net learning function updates weight and bias values conform to Levenberd-Marquerdet optimization algorithm (Hagan and Menhaj, 1994).

In order to assess the individual contribution of input variables to the prediction process of ANN, a sensitivity technique was tested. The technique was based on disabling, during all phases (training, evaluating, testing and simulation), one processing element (neuron) in the input layer and comparing those results with the standard ANN (1 and 2).

The criteria used to compare the results of ANN1 and ANN2 anticipation with the actual observed data were: Pearson coefficient of correlation between observed and predicted value, root mean square error and Theil inequality coefficient and a correlated (samples) t-test was applied for comparison difference between observed and predicted means. The test was performed by the t-test procedure in SAS (1997).

$$r_p = \frac{\delta_{ip}}{\delta_i \delta_p} \quad (1)$$

Where, r_p = Pearson correlation coefficient between observed and predict data, δ_{ip} = covariance between observed and predicted data, δ_i = standard deviation of observed data, δ_p = standard deviation of predicted data.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2}{n}} \quad (2)$$

Table 1. Data structure for the observed and predicted (ANN1 and ANN2) data.

Data		Min	Max	Mean	SD	CV
OBS	sc	235.00	410.00	303.75	68.73	22.62
	tsc	117.50	575.00	295.13	160.36	54.33
ANN1		250.2	397.86	297.28	50.04	16.38
ANN2		164.15	583.46	285.48	151.93	53.22

OBS: Observed value; Sc: semen concentration in 100 cc volume; tsc: total semen production in total volume; ANN1: artificial neural network prediction for the sc; ANN2: artificial neural network prediction for the tsc.

Table 2. Correlation and comparisons between observed and predicted (ANN1 and ANN2) data.

Data	t	r_p	RMSE	%RMSE	I^2
ANN1	0.22^{ns}	0.82***	37.41	12.58	0.01453
ANN2	0.12^{ns}	0.94***	52.79	18.49	0.02543

*** p <0.001; ns: p > 0.05; OBS: Observed value; ANN1: artificial neural network prediction for the sc; ANN2: artificial neural network prediction for the tsc; t: t value for mean difference between the observed and predicted data; %RMSE: RMSE divided by the mean of performance.

Where, RMSE = Root mean square error, n = number of records, y_i = observed value, $\hat{y}$ = estimated values by ANN_1 and ANN_2.

$$I^2 = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2}{\sum_{i=1}^{n} y_i^2} \quad 3$$

Where, I^2 is Theil inequality coefficient (Theil, 1979) and the other symbol are the same as previous formulas.

RESULTS AND DISCUSSION

The variability parameter in Table 1 (SD, CV) was also closer to those of the observed data for ANN1 than ANN2. Table 2 results showed that there was no significant difference between ANN1 and ANN2 predictions and the observed values (p > 0.05). This shows that the results of the two ANNs are reliable for both criteria of semen production trait. The high correlations showed that the predicted averages for semen concentration were close to the observed values (Table 2). Thus, ANNs are reliable decision support system that helps breeders choose rams to be left out or culled from the herd.

Lower RMSE and coefficient correlation of (r_p) for ANN1 and ANN2 showed antagonism in results by concentration criteria. However, ANN1 r_p is lower than ANN2 but ANN2 have more prediction error (Table 2) although I^2 criteria which enclosed the ANN1 have better network structures for prediction. This result may be as an effect of range of data in output layer (tsc have extreme range and higher SD than SC criteria in all category). Moreover, modification of learning or training parameter and the method of data presentation can considerably influence the network performance (Salehi et al., 1998). The performance of ANN2 with tsc data seems to be more justified because it has a lower correlation coefficient which does not correspond to the range of data. Prediction accuracy (r_p) increases by tsc data in relation to the use of Sc data in output layer. This may be related to the training, proper update weights and less bias in the network system as a result of high correlation coefficients with the observed value and the fact that as good as the well-structured presented data of ANN is, there should be a better training for ANN such as orthogonizing and classification of input vector to provide better predictions and improve the predictive ability of ANNs due to the fact that they perform particularly well in interpolation (Lacroix et al., 1995).

As the training concept suggests, prediction of unseen records by a neural network improves when similar cases have been included during training (Salehi et al., 1998). Thus, selection of a proper sample is very important for training on ANN. The sample data must be a reflection of all events in population proper rather than proportion event, and should be considerable in effect of range of data in input and output layer on network response. Pre-processing data (e.g., standardization and normalization) may lead to an improvement in the learning process of ANNs (Stein, 1993) which helps neural networks to predict better.

Sensitive analysis of the ANN model (Table 3) allowed us to stress the fundamental importance of the variable "testis circumference", which denotes a positive correlation coefficient with semen concentration (Sablani et al., 1995) but testis circumference with weight of testis has a less correlation, and this may be as a result of different

Table 3. Coefficient and RMSE value obtained with ANNs after disabling one processing element in tow volume of sc and tsc.

Missing variable	volume	r_p	RMSE	%RMSE
testis volume	Sc	0.75	84.18	32.74
	tsc	0.88	42.61	14.76
testis dimension	Sc	0.83	58.46	19.57
	Tsc	0.91	80.92	35.37
testis circumference	Sc	0.40	42.61	14.76
	tsc	0.77	97.95	34.57

testis wool cover, fat under skin and high testis skin. The "testis volume" has a lower importance than testis circumference. Testis volume trait has linear relationship with weight of body, and the highest rams which have the biggest testis, produce more semen (Sablani et al., 1995). Finally, "testis dimension" has lower effect in prediction of semen concentration. However, the seasons have magnitude effect on that entire trait.

Conclusion

The major use of any predictive process is to support accurate decisions which are dependent on prior knowledge of the possible outcome. The outcome of this study showed that, generally, artificial neural network have the potential to play an important role in modeling biological processes and there are many potential application areas in predicting semen production with high accuracy. The result showed that phenotypic trait information can be used in the prediction of product of semen concentration.

The accuracy of ANN will be more improved when variables which are more relevant to the output variables are used. Once neural network is trained and has been shown to be effective, it will be easy to use. Also, ANN has a good potential to be used in the prediction of future records of rams for setup selection program. This increases the genetic potential of sheep herds and it is a good support system for farmer for decision making. With regard to this, earlier and more accurate prediction of semen production should have a beneficial effect on the timely identification of genetically superior animals.

REFRENCES

Adamczyk K, Molenda K, Szarek J, Skrzyński G (2005). Prediction of bulls slaughter from growth data using artificial neural network. J. Cent. Eur. Agric. 6: 133-142.

Blache D, Chagas LM, Blackberry1 MA, Vercoe PE, Martin GB (2000). Metabolic factors affecting the reproductive axis in male sheep. J. Reprod. Fertil. 120: 1-11.

Hagan MT, Menhaj M (1994). Training feed forward networks with the Marquardt algorithm. IEEE Trans. Neural Networks. 5: 989-993.

Hosseinnia P, Edrisi M, Edriss MA, Nilforooshan MA (2007). Prediction of Second Parity Milk Yield and Fat Percentage of Dairy Cows Based on First Parity Information Using Neural Network System. J. Appl. Sci. 21: 3274-3279

Kominakis AP, Abas Z, Maltaris I, Rogdakis E (2002). Preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. Comput. Electron. Agric. 1: 35-48

Lacroix R, Salehi F, Yang XZ, Wade KM (1997). Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. Trans. ASAE, 40: 839-846

Lacroix R, Wade KM, Kok R, Hayes JF (1995). Prediction of cow performance with a connectionist model. Trans. ASAE, 38: 1573-1579.

MATLAB (2006). MATLAB Guide for Personal Computers. Version 7.3 ed. MATLAB Institute, USA.

Ruan R, Almaer S, Zou C, Chen PL (1997). Spectrum analysis of mixing power curves for neural networks prediction of dough rheological properties. Trans. ASAE, 40: 677-681.

Sablani SS, Ramaswamy HS, Prasher SO (1995). A neural network approach for thermal processing applications. J. Food Proc. Preserv. 19: 283-301.

Salehi F, Lacroix R, Wade KM (1998a). Improving dairy yield through combined record classifiers and specialized artificial neural network. Comput. Electron. Agric. 20: 199-213.

Salehi F, Lacroix R, Wade KM (1998b). Effects of learning parameters and data presentation on the performance of back-propagation networks for milk yield prediction. Trans. ASAE, 41: 253-259.

SAS (1997). User's Guide: Statistics, Version 6.12 Edition. SAS Inst., Inc., Cary, NC.

Stein R (1993). Selecting data for neural networks. AI Expert. 2: 42-47

Theil H (1979). World income inequality. Econ. Lett. 2: 99-102.

Wasserman PD (1993). Advance Method in Neural Computing. Van Nostrand Reinhod, New York, p. 225.

Yang XZ, Lacroix R, Wade KM (1999). Neural detection of mastitis from dairy herd improvement records. Trans. ASAE, 42: 1063-1072.