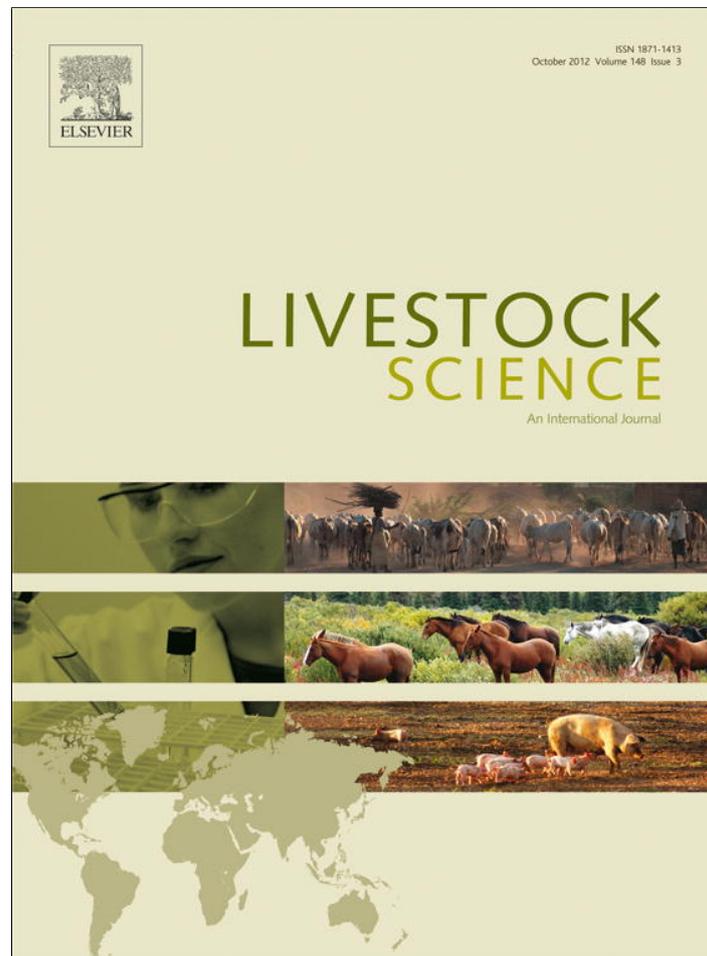


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

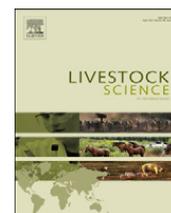
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

## Livestock Science

journal homepage: [www.elsevier.com/locate/livsci](http://www.elsevier.com/locate/livsci)

# A neural network model to describe weight gain of sheep from genes polymorphism, birth weight and birth type

Mojtaba Tahmoorespur, Hamed Ahmadi\*

Animal Sciences Department, Ferdowsi University of Mashhad, P.O. Box 91775-1163, Iran

## ARTICLE INFO

## Article history:

Received 30 January 2012

Received in revised form

20 April 2012

Accepted 4 June 2012

## Keywords:

Artificial neural network

Growth traits

Polymorphism

Baluchi sheep

## ABSTRACT

Polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP) method was used to determine the growth hormone (GH), leptin, calpain, and calpastatin polymorphism in Iranian Baluchi male sheep. An artificial neural network (ANN) model was developed to describe average daily gain (ADG) in lambs from input parameters of GH, leptin, calpain, and calpastatin polymorphism, birth weight, and birth type. The fitness of the model was tested using  $R^2$ , MS error, and bias. The developed ANN-model was used to evaluate the relative importance of each input parameter on lambs ADG using a sensitivity analysis method. Three conformational patterns were detected for GH, Leptin, calpain genes, and five conformational patterns were detected for calpastatin gene. The calculated statistical values corresponding to the ANN-model showed a high accuracy of prediction ( $R^2 > 0.90$ , MS error=0.0003). The sensitivity analysis on the ANN-model indicated that birth weight and birth type are the most important variables in the growth of lambs, followed by Leptin, calpastatin, GH, and calpain, polymorphism. The optimization analysis on ANN-model for maximizing ADG of lambs revealed that the maximum ADG may be obtained with birth weight 5.2 kg, birth type of single, GH genotype of G2, leptin genotype of L1, calpastatin genotype of C3, and calpain genotype of A3. The results revealed that the ANN-model is an appropriate tool to recognize the patterns of data to predict lamb growth in terms of ADG given specific genes polymorphism, birth weight, and birth type. The platform of PCR-SSCP approach and ANN-based model analyses may be used in molecular marker-assisted selection and breeding programs to design a scheme in enhancing the efficacy of sheep production.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

The lamb meat has a large contribution to supply human protein consumption in Iran. However, the sheep production is not sufficient to cover the consumers demand. An appropriate breeding scheme may help to enhance the efficacy of sheep production (Yazdi et al., 1997). Effective sheep production for meat depends on lamb weight, thus, breeding objectives should concentrate on this trait (Tosh and Kemp, 1994). Quantitative traits such as average daily gain (ADG) or

final body weight are controlled by a great number of genes. Several genes such as growth hormone (GH), leptin, calpain, and calpastatin are considered as genetic candidate markers for growth rate and meat production traits in mammalian (Chung et al., 2007; Marques et al., 2006; Nassiry et al., 2007; Zhou et al., 2009). Investigation into the relationship between these genetic markers polymorphism and ADG of lambs may lead to design a more efficient selection program. In this way, statistical and mathematical tools such as linear regression, logistic regression, and mixed models may be used. Alternatively, a soft-computing method of artificial neural network (ANN) has shown a great ability for solving complex nonlinear system identification and control problems. The ANN process is a method of data analysis that is assumed to imitate the human brain's way of working (i.e. training and

\* Corresponding author. Tel.: +98 913 273 5084;

fax: +98 511 879 6845.

E-mail address: hahmadima@yahoo.com (H. Ahmadi).

memory). The ANN-model is an interconnected group of artificial neurons that process information in a parallel form. The information is stored in the network's "synaptic" connections (Dayhoff and DeLeo, 2001). There are three kinds of layers in ANN: input layer, one or more hidden layers, and output layer. The neurons are connected with each other with a different connection strength. The connections are called synaptic weights. In weights, the whole information on the network is encoded because those weights are in fact the numbers that determine the strength of the stimuli coming to the neurons. The most important feature of ANN, which determines the specificity of that computational method, is the process of training. Training of the networks is realized by the changes of the values for all synaptic weights with the use of a specific algorithm (Haykin, 1999). The most commonly used learning algorithm is the so-called back-propagation training algorithm. In the process of learning with that algorithm, the network uses the error between the actual (observed) and the desirable output to improve the values of synaptic weights (Dayhoff and DeLeo, 2001). Recently, the ANN has been applied in many fields to model and predict the behaviors of systems, based on given input–output data. Successful applications of ANN have been reported in various animal subjects (Ahmadi and Golian, 2010a, b; Gianola et al., 2011; Grzesiak et al., 2006; Heald et al., 2000; Salehi et al., 1998). The ANN-base modeling has gained much attention; however, the use of such method is not common in animal genetic field.

The objectives of this study were: (1) to detect GH, leptin, calpain, calpastatin polymorphism in Iranian male Baluchi sheep; (2) to develop an ANN-based model to describe ADG of lambs during 0–3 months of age based on investigating genes polymorphism, birth weight, and birth type; (3) to apply the developed ANN-model to evaluate the relative importance of the genes, birth weight, and birth type on ADG of lambs; and (4) to optimize ANN-model to find the optimum values of input variable for maximizing ADG of lambs.

## 2. Materials and methods

### 2.1. Animals and polymorphism detection

Seventy-five purebred male Baluchi sheep from the Baluchi Breeding Station located in north-east of Mashhad, Iran were used in this study. Lambs were weighed and ear-tagged at birth time. Weaning was done at 3

months of age. The lambs were weighed monthly and the ADG was calculated using birth weight and 3-month weight values. Individual blood samples were obtained from jugular vein using disposable syringes into sample tubes containing the anticoagulant ethylenediaminetetraacetic acid 10% (EDTA). Genomic DNA was extracted from the samples using the method by Boom et al. (1990) with minor modifications. After measuring the DNA concentration and its purity by Nano-Drope ND 2000 spectrophotometer (Wilmington, Delaware, USA), DNA was diluted to a final concentration of 50 ng/μL in dH<sub>2</sub>O and stored at 4 °C. All primer sets for ovine GH, leptin, calpain, and calpastatin genes were amplified in separate reactions and primers sequence is shown in Table 1. The PCR amplification was carried out in a final volume of 25 μL containing 75 mM Tris–HCl (pH 8.8), 1 unit of Platinum Taq DNA Polymerase (Invitrogen, USA), 0.2 mM each of dATP, dCTP, dGTP, dTTP (Pharmacia, Uppsala, Sweden), 1.5 mM MgCl<sub>2</sub>, 10 pmol of primers and 50 ng of DNA template. Gene's polymorphisms were detected using single-strand conformation polymorphism (SSCP). More details of DNA amplification and SSCP analysis is described by Tahmoorespur et al. (2011).

### 2.2. Data preprocessing

The detected polymorphism and birth type data as ANN-model inputs (categorical inputs) were converted to numerical values of 101–105 (Table 2). Prior to training, the data set (input and output data) were normalized within the range [–1, 1]. This is to simplify the problem for the network, to achieve fast convergence minimum mean square error (MSE), and to ensure the fall of targets (output data) into the specific range that the new feed forward network can be reproduced.

### 2.3. Artificial neural network model development and evaluation

A feed-forward multi-layer perceptron as a most common network structure was used to construct the ANN-model. It was trained by the back-propagation algorithm and the conjugate gradient method. The variables of interest in this multi-input, single-output model were:

Categorical inputs ( $n=8$ ): The GH, leptin, calpain, calpastatin polymorphism, and birth type.

Continues input ( $n=1$ ): birth weight.

**Table 1**  
Single-strand conformation polymorphism primers for ovine genes.

Genes <sup>1</sup>	Annealing temperature (°C)	Product length (bp)	Primer sequence
Growth hormone	58.0	366	Forward 5'-GAAACCTCCTCTCGCCC-3' Reverse 5'-CCAGGGTCTAGGAAGGCACA-3'
Leptin	66.5	275	Forward 5'-GCTCCACCCTCTCTGAGTTTGTC-3' Reverse 5'-GTCTGTAGAGACCCCTGTAGCCG-3'
Capastatin	61.0	622	Forward 5'-TGGGGCCCAATGACGCCATCGATG-3' Reverse 5'-GGTGGAGCAGCACTTCTGATCACC-3'
Calpain	58.0	190	Forward 5'-AACATTCTCAACAAAGTGGTG-3' Reverse 5'-ACATCCATTACGCCACCAT-3'

**Table 2**

Characteristics of detected genes polymorphism and conversion values of detected genotype into the numerical values.

Gene name	Number of detected conformational patterns (genotype)	Conversion detected genotype into numerical value
Growth hormone	3	G1=101, G2=102, G3=103
Leptin	3	L1=101, L2=102, L3=103
Calpastatin	5	C1=101, C2=102, C3=103, C4=104, C5=105
Calpain	3	A1=101, A2=102, A3=103

Continues output ( $n=1$ ): The ADG of lambs during 0–3 months of age.

The 75 data lines were randomly divided into 2 sets of learning and testing with 55 and 20 data lines, respectively (Table 1 in Supplemental file). The data set was imported into the Statistica Neural Networks software version 7.1 (StatSoft, 2005). A common problem in ANN training is over fitting. A network with a large number of weights in comparison with the number of training cases available may achieve a low training error by modeling a function that fits the training data well. An over-fitted model typically has high curvature, as the function is contorted to pass through the points, modeling any noise in addition to the underlying data (StatSoft, 2005). In this study, the optimum architecture of network was determined using the selection algorithms integrated in the “intelligent problem solver” module of Statistica software. To avoid over fitting problem, a weigend regularization as a modification to the error function was used. In this way, the large weight values are penalized, allowing a network to automatically determine its own level of complexity and so avoid over fitting (Weigend et al., 1991). This leads the ANN process to develop a smoother model. Thus, the optimum architecture of the perceptron neural network (i.e., the number of hidden neurons in the network and training algorithm) was determined using the selection algorithms integrated in the “intelligent problem solver” module of Statistica software. A random group of values with normal distribution was automatically produced and would be used to set the initial weight during the training process. The goodness-of-fit of the ANN-model was determined using  $R^2$ , MS error, and bias (Ahmadi et al., 2007).

#### 2.4. Sensitivity analysis

The sensitivity analysis indicates which input variable is considered most important by the ANN-model. The sensitivity is defined as the ratio between the error with omission and the baseline error and ranks the variables in the order of importance (Hunter et al., 2000). The sensitivity of an input variable against the output variable(s) may be determined using following criteria (Lou and Nakai, 2001; StatSoft., 2005).

The variable sensitivity error (VSE) value indicates the performance of the developed model if that variable is unavailable.

The variable sensitivity ratio (VSR) value is a relative indication of the ratio between the VSE and the error of the model when all variables are available.

A more important variable has a higher VSR value. Thus, according to the obtained VSR value, the input variables could be ranked in the order of importance (Ahmadi and Golian, 2010a, b).

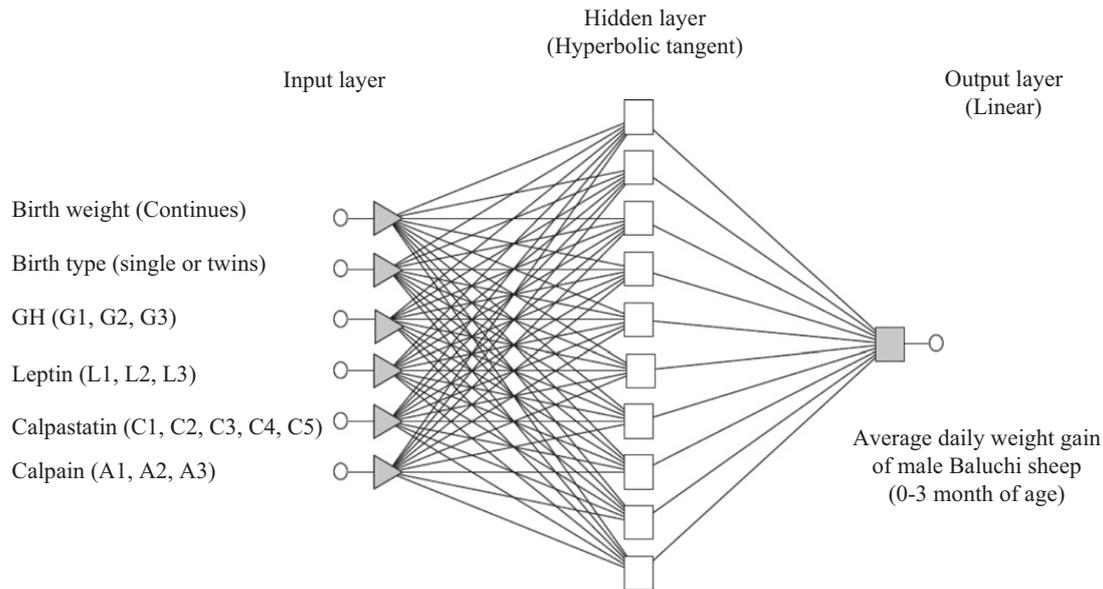
#### 2.5. Model optimization

The constructed ANN-model was subjected to a further process of optimization to find the optimal values of input variables for maximizing ADG of lambs. A “random search” optimization algorithm provided in the “response optimization for data mining models” section of Statistica software was used (StatSoft., 2005). This linear random search technique is unguided algorithms based on brute computing power. The iterative optimization algorithm takes samples from the space of the input variable values. For each values sample, the model prediction is computed and compared with the best value found from the previous iterations. If the newly found value is better than the previous one, the new results are stored. This process is repeated until the end of iterations is reached (Ahmadi and Golian, 2010a).

### 3. Results and discussion

In an elementary phase of this research, PCR-SSCP results indicated that three conformational patterns were detected for GH (genotypes G1, G2 and G3), Leptin (genotypes L1, L2 and L3), calpain (genotypes A1, A2 and A3) genes; and five conformational patterns were detected for calpastatin (genotypes C1, C2, C3, C4 and C5) gene. In the last decade, biomolecular techniques such as PCR have received particular attention. More recently, DNA markers such as SSCP is more widely applicable and an appropriate tools for evaluating genetic variation for quantitative traits and can readily be used to assist in genetic selection of superior animals. Several authors in the last decade reported that SSCP analysis is a reliable and repeatable technique for detection of gene polymorphism due to point mutations (Nassiry et al., 2007; Tahmoorespur et al., 2011). However, the SSCP technique is recently being supplanted by sequencing, so the high throughput genotyping technology has proved to be more efficient than quantitative trait loci and marker-assisted selection strategies.

The optimal architecture of the multi-layer perceptron neural network for modeling ADG that was suggested by “intelligent problem solver” was found with 6 inputs, one output (with linear activation function), and 10 hidden neurons (with hyperbolic tangent activation function). A training algorithm of Quasi-Newton was used to train the network (Lou and Nakai, 2001; StatSoft., 2005). The general scheme of such neural network model structure is outlined in Fig. 1.



**Fig. 1.** Architecture of the three-layer perceptron neural network used for modeling average daily gain of male Baluchi sheep consisting: one continues input of birth weight and eight categorical inputs of growth hormone (GH), leptin, calpain, calpastatin polymorphism and birth type; one continues output of average daily gain of male Baluchi sheep (from 0 to 3 months of age); 10 hidden neurons with hyperbolic tangent activation function. A training algorithm of Quasi-Newton was used to train the network.

The optimization of the architecture by the intelligent problem solver was performed on the basis of the “balance error against diversity” option. This option tries to produce a structure with a balance performance value against type and diversity (StatSoft., 2005). It will preserve networks with a range of types and performance/complexity trade-offs. This may led to obtain an optimized ANN-model with less complexity and more accuracy. The predicted values of ADG by the ANN-model are shown in Supplemental file (Table 1). The comparison of observed and predicted outputs describes the behavior of the ANN-model from investigating inputs. The results revealed a good agreement between the observed and the predicted values of ADG for training and testing sets (Fig. 2). The calculated statistics on the ANN-model are in close agreement to the two subsets in prediction of each output ( $R^2=0.91$  and  $0.92$  for training and testing sets, respectively;  $MSE=0.0003$  for both training and testing sets). A well trained ANN-model has a balance statistics values for these two subsets. This may suggest that over fitting has not occurred during the training process (Ahmadi and Golian, 2010b; Dayhoff and DeLeo, 2001; Lou and Nakai, 2001). The main advantage of ANN-model is that it does not require a prior specification of suitable fitting function thus, it has a universal approximation ability to approximate almost all kinds of nonlinear functions. This flexibility feature may help the modeler to make a model with almost highest possible prediction accuracy. The bias value implies the magnitude of the model over/under estimation regarding to the average of observed values. Calculated bias values for current ANN-model showed no (bias=0.0000) over/under fitting by the developed model both in training and testing sets.

To determine the relative importance of input variables, the entire 75 data lines (training and testing)

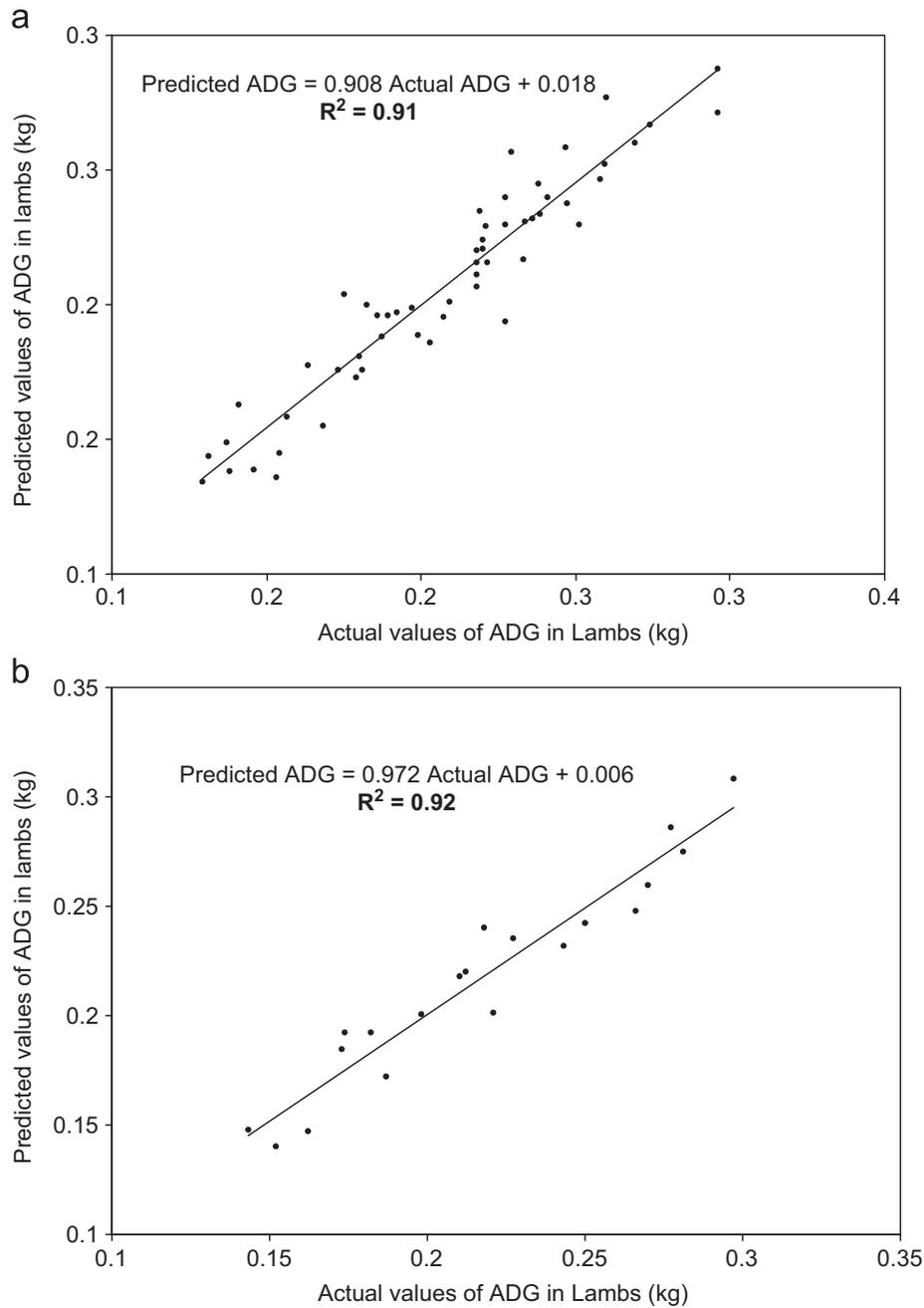
utilized to calculate the overall VSE and VSR. The obtained VSR for the model output variable (ADG) with respect to GH, leptin, calpain, calpastatin polymorphism, birth type, and birth weight are shown in Table 3.

A more important input variable has a higher VSR value. Thus, the investigating inputs were ranked according to their importance of effect on ADG using VSR values. Among the input variables, birth weight (2.68) and birth type (2.53) have the highest values of VSR in data set. They are followed by leptin (2.5), calpastatin (2.08), GH (2.03), and calpain (1.48).

The optimization analysis on ANN-model for maximizing ADG of lambs revealed that the maximum ADG may be obtained with birth weight 5.2 kg, birth type of single, GH genotype of G2, leptin genotype of L1, calpastatin genotype of C3, and calpain genotype of A3. At the optimum point, the predicted ADG of lambs was 295 g/lamb/d. To enhance the efficacy of Iranian Baluchi sheep production (in term of ADG) the optimization results suggest that focusing on this genotype pattern may help to develop a more efficient breeding program.

#### 4. Conclusions

The SSCP approach was used to detect GH, leptin, calpain, calpastatin genes in Iranian Baluchi sheep. The ANN-model was successfully used to describe associations between investigating gene polymorphism, birth weight and birth type and ADG of sheep. The sensitivity analysis on the ANN-model indicated that birth weight and birth type is the most important variable in the growth of lambs, followed by leptin, calpastatin, GH, and calpain genes polymorphism. The results suggested the possibility of using the SSCP approach and ANN-base model analyses in molecular marker-assisted selection



**Fig. 2.** The scatter plot of actual vs. model predicted values for average daily gain (ADG) of male Baluchi sheep from 0 to 3 months of age (black points): (a) training set ( $n=55$ ) and (b) testing set ( $n=20$ ). Solid line indicating the fitted simple regression line on scattered points.

**Table 3**

Calculated variable sensitivity ratio (VSR) for the global sensitivity analysis of input variables in the neural network model for average gain (from 0 to 3 months of age) in male Bluchi sheep.

	Input variables					
	Birth weight	Birth type (single or twins)	Growth hormone	Leptin	Calpastatin	Calpain
VSR	2.68	2.53	2.03	2.50	2.08	1.48
Ranking based on importance	1	2	5	3	4	6

and breeding programs. The genotype scheme generated by optimization of ANN-model in maximizing ADG of lambs may be considered as an applied result for experts in practical situation to develop an appropriate breeding

plan to enhance the efficacy of sheep production. However, additional research with a larger population of the same breed is required to obtain a more appropriate and robust model.

**Conflict of interest statement**

None.

**Acknowledgments**

This study was financially supported by Ferdowsi University of Mashhad, affair of research and technology; Experiments were conducted in Animal Biotechnology Labratory at Ferdowsi University of Mashhad.

**Appendix A. Supplementary Information**

**Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.livsci.2012.06.008>.**

**References**

- Ahmadi, H., Golian, A., 2010a. Growth analysis of chickens fed diets varying in the percentage of metabolizable energy provided by protein, fat, and carbohydrate through artificial neural network. *Poult. Sci.* 89, 173–179.
- Ahmadi, H., Golian, A., 2010b. The integration of broiler chicken threonine responses data into neural network models. *Poult. Sci.* 89, 2535–2541.
- Ahmadi, H., Mottaghtalab, M., Nariman-Zadeh, N., 2007. Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. *J. Appl. Poult. Res.* 16, 494–501.
- Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheimvanden, P.M.E., Vandernoordaa, J., 1990. Rapid and simple method for purification of nucleic-acids. *J. Clin. Microbiol.* 28, 495–503.
- Chung, H., Choi, B., Jang, G., Lee, K., Kim, H., Yoon, S., Im, S., Davis, M., Hines, H., 2007. Effect of variants in the ovine skeletal-muscle-specific calpain gene on body weight. *J. Appl. Genet.* 48, 63–68.
- Dayhoff, J.E., DeLeo, J.M., 2001. Artificial neural networks—opening the black box. *Cancer* 91, 1615–1635.
- Gianola, D., Okut, H., Weigel, K.A., Rosa, G.J., 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genet.* 7, 12–87.
- Grzesiak, W., Blaszczyk, P., 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. Agric.* 54, 69–83.
- Haykin, S., 1999. In: *Neural Networks: A Comprehensive Foundation* 2nd ed. Prentice-Hall, Upper Saddle River, NJ.
- Heald, C.W., Kim, T., Sischo, W.M., Cooper, J.B., Wolfgang, D.R., 2000. A computerized mastitis decision aid using farm-based records: an artificial neural network approach. *J. Dairy Sci.* 83, 711–720.
- Hunter, A., Kennedy, L., Henry, J., Ferguson, I., 2000. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Comput. Methods Programs Biomed.* 62, 11–19.
- Lou, W., Nakai, S., 2001. Artificial neural network-based predictive model for bacterial growth in a simulated medium of modified-atmosphere-packed cooked meat products. *J. Agric. Food Chem.* 49, 1799–1804.
- Marques, M.D., Santos, I.C., Carolino, N., Belo, C.C., Renaville, R., Cravador, A., 2006. Effects of genetic polymorphisms at the growth hormone gene on milk yield in Serra da Estrela sheep. *J. Dairy Res.* 73, 394–405.
- Nassiry, M.R., Shahroudi, F.E., Tahmoorespur, M., Javadmanesh, A., 2007. Genetic variability and population structure in beta-lactoglobulin, calpastain and calpain loci in Iranian Kurdi sheep. *Pak. J. Biol. Sci.* 10, 1062–1067.
- Salehi, F., Lacroix, R., Wade, K.M., 1998. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Comput. Electron. Agric.* 20, 199–213.
- StatSoft., 2005. *Statistica. Data Analysis Software System.* StatSoft Inc., Tulsa, OK.
- Tahmoorespur, M., Taheri, A., Gholami, H., Ansary, M., 2011. PCR-SSCP variation of GH and STAT5A genes and their association with estimated breeding values of growth traits in Baluchi Sheep. *Anim. Biotechnol.* 22, 37–43.
- Tosh, J.J., Kemp, R.A., 1994. Estimation of variance-components for lamb weights in 3 sheep populations. *J. Anim. Sci.* 72, 1184–1190.
- Weigend, A.S., Rumelhart, D.E., Huberman, B.A., 1991. Generalization by weight-elimination with application to forecasting. In: Lippmann, R.P., Moody, J.E., Touretzky, D.S. (Eds.), *Advances in Neural Information Processing Systems*, 3. , Morgan Kaufmann, San Mateo, CA, pp. 875–882.
- Yazdi, M.H., Engstrom, G., Nasholm, A., Johansson, K., Jorjani, H., Liljedahl, L.E., 1997. Genetic parameters for lamb weight at different ages and wool production in Baluchi sheep. *Anim. Sci.* 65, 247–255.
- Zhou, H., Hickford, J.G.H., Gong, H., 2009. Identification of allelic polymorphism in the Ovine Leptin Gene. *Mol. Biotechnol.* 41, 22–25.