Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
29 (3): 289-292, 2023
DOI: 10.9775/kvfd.2022.28672

Short Communication

The Use of Artificial Neural Networks for Prediction of Milk Productivity of Cows in Ukraine

Mykhailo MATVIEIEV [1] Yuriy ROMASEVYCH [2] Andriy GETYA [3 (*)]

[1] National University of Life and Environmental Sciences of Ukraine, Faculty of Livestock Raising and Water Bioresources, Department of Animal Feeding and Feed Technology named after Pshenychnyi, 03041 Kyiv, UKRAINE

[2] National University of Life and Environmental Sciences of Ukraine, Faculty of Design and Engineering, Department of Machines and Equipment Design, 03041 Kyiv, UKRAINE

[3] National University of Life and Environmental Sciences of Ukraine, Faculty of Livestock Raising and Water Bioresources, Department of Animal Genetics, Breeding and Animal Biotechnology, 03041 Kyiv, UKRAINE

ORCIDs: M.M. 0000-0003-1281-9032; Y.R. 0000-0001-5069-5929; A.G. 0000-0002-4747-9261

Article ID: KVFD-2022-28672 Received: 24.10.2022 Accepted: 06.04.2023 Published Online: 12.04.2023

Abstract: Aim of this work was to study effectiveness of prediction of cows' milk productivity (yield, fat and protein content) using artificial neural network (ANN) technology for data sets with missing values. Four variants of data sets, which consist of different numbers of monthly test-day milk records (MTDMRs) were chosen. Calculated milk productivity using Test Interval Method (TIM) was chosen as control value. Obtained results showed that milk productivity can be predicted using ANN even if missing data occurs. Rank correlation coefficients between control and predicted results were 0.918, 0.949, and 0.852 for milk yield, protein, and fat content, respectively.

Keywords: *Artificial Neural Networks, Dairy Productivity Prediction*

Ukrayna'da İneklerin Süt Verimliliğinin Tahmininde Yapay Sinir Ağlarının Kullanımı

Öz: Bu çalışmanın amacı, ineklerin süt verimliliğinin (verim, yağ ve protein içeriği) tahmin edilmesinde eksik değerlere ait veri setleri için yapay sinir ağı (YSA) teknolojisini kullanılarak etkinliğini incelemektir. Farklı sayıda aylık test-günlük süt kayıtlarından (MTDMR) oluşan dört farklı veri seti seçilmiştir. Test Aralık Yöntemi (TIM) kullanılarak hesaplanan süt verimliliği kontrol değeri olarak kullanılmıştır. Elde edilen sonuçlar, eksik veriler olsa bile süt verimliliğinin YSA kullanılarak tahmin edilebileceğini göstermiştir. Kontrol ile tahmin edilen sonuçları arasındaki sıralama korelasyon katsayıları, süt verimi, protein ve yağ içeriği için sırasıyla 0.918, 0.949 ve 0.852 olarak saptanmıştır.

Anahtar sözcükler: *Yapay Sinir Ağları, Süt Verimliliği Tahmini*

Introduction

The improvement of modern breeds of farm animals occurs with the widespread use of mathematical methods integrated into software. One of the basic elements of successful application of programs is the availability of data on the productivity of animals, collected as a result of organized reliable breeding [1]. The list of traits is usually regulated by national and international associations and unions. Maintenance of such recording scheme is important for both, traditional breeding and genomic [2].

In the world, there is a tendency for a constant increase in the number of traits, particularly in dairy farming [3], which necessitated the use of modern methods of data processing, including ANN, machine learning algorithms, and other methods [4].

In Ukraine ANN did not find their application in breeding practice. However, the main problem in Ukraine is not a large amount of data, but the presence of incomplete data (missing values) in productivity records, as in many farms it is not recorded regularly.

How to cite this article?
Matvieiev M, Romasevych Y, Getya A: The use of artificial neural networks for prediction of milk productivity of cows in Ukraine. *Kafkas Univ Vet Fak Derg,* 29 (3): 289-292, 2023.
DOI: 10.9775/kvfd.2022.28672

(*) Corresponding author: Andriy GETYA
Phone: +38 050 1950125 Fax: +38 050 1950125
E-mail: getya@ukr.net

Therefore, it is often not possible to assess milk productivity by classical methods. In this case, there are several ways to predict milk productivity of cows [5].

Based on the above mentioned, it was decided to study the effectiveness of prediction of milk productivity of cows using the capabilities of ANN technology for the cases with missing values about milk productivity.

Material and Methods

Ethical Statement

This study was approved by the Commission on bioethics of Research Institute of Animal Health of the National University of Life and Environmental Sciences of Ukraine (Approval no: 103-07 from 01.07.2019).

General Conditions

Cows of different origins, years of birth, and lactation were involved in the study. Their milk yield exceeded 6000 kg per standard lactation (lactation duration 305 days). Analysis of milk, collected on the farm, was performed in the laboratory. Contents of fat (%), protein (%), lactose (%), and somatic cell counts (SCC, thousand/cm^3) were determined [6].

To transform the somatic cell counts (SCC) into the somatic cell scores (SCS) the formula by Wiggans and Shook [7] was applied.

For prediction of cows' productivity 4 variants of the data sets were tested, which included different MTDMRs counted from the beginning of lactation:

1) first, second, fifth, eighth, and tenth MTDMRs; 2) first three, ninth and the tenth MTDMRs; 3) first five MTDMRs; 4) second, fifth, and tenth MTDMRs.

Obtained results were compared with results, calculated according to TIM [8], using the data of 10 MTDMRs of 144 cows.

The animals whose data were included in the prediction of productivity were the same for all four variants.

Development of ANNs and Their Training

Procedures for training ANNs and prediction were carried out for each data set separately. For training of each ANN, milk productivity data, calculated using TIM on the basis on 10 MTDMRs, were used.

A feedforward ANN was used for training. The number of inputs of each ANN corresponded to the dimension of the input vector for a certain task (36 - for ANN-1, ANN-2 and ANN-3, 24 - for ANN-4). For all tasks, the output of the ANN was represented by a four-component vector of the following traits (data): milk yield, fat and protein content per standard lactation.

The number of hidden layers of the ANN was equal to one, and the number of neurons in the hidden layer was 10. The activation function for all neurons is a hyperbolic tangent. For all of the neurons biases were used as well.

The estimation of unknown ANNs' parameters corresponds to the total number of biases and weights. The number of biases in the hidden layer equals 10 and in the output layer - 4. The ANNs' weights form two matrices: the first one for the hidden layer, and the second one - for the output layer. For the ANN-1, ANN-2, and ANN-3 cases, the hidden layer weights matrix is of 36x10 dimension. For the ANN-4 the matrix is of 24x10 dimension. The output layer weights matrix for all of the ANNs is of 10x4 dimension. This data allows the determination of a total number of needed parameters: 414 weights and biases for ANN-1, ANN-2, and ANN-3, 294 weights and biases for ANN-4.

The function that reflected the quality of training (loss function) was the root mean square deviation of the prediction and training data. Moving on the surface of the loss function a training algorithm must determine such values of weights and biases, that the corresponding value of the loss function reaches a minimum (the best variant - the global one). This problem is quite difficult, because of the huge dimensionality of the loss function, and its topology complexity (non-linearity, multimodality, non-separability, etc.). Thus, the selection of a training algorithm is a very important stage of the study, it influences ANN prediction quality. Here we applied the gradient-base ADAM algorithm [9], which is a common and effective method of ANNs training. In the ADAM algorithm, the size of the data sets, on which one gradient was calculated, was chosen to be 50. The number of training rounds was 5000 (this number was defined to be sufficient to minimize the loss function).

The training pair in the prediction tasks described above was represented by data in the following format: input vector (traits of the cow) - scalar (productivity of the cow per standard lactation). The input vector included a different number of components for different task statements.

In all samples, the individual cow number, year of birth of the cow, sire number (coded with numbers from 1 to 51), calving age (in days from birth of cow), day of MTDMR (in days from calving) and performance data for each MTDMR (milk yield, fat-, protein-, lactose content, and SCS) were considered.

All data were normalized. So, numerical values in both groups (for training and for prediction) for all traits vary from 0 to 1. This makes it possible to train ANNs since input vector components vary in the limited domains, and no activation function saturation will occur. The feature of the current work, which should be stressed, is connected with quite a big number of traits in data. This provides additional factors to involve in prediction. Their influence we might assess indirectly by analyzing the prediction performance of a trained ANN.

Verification of the Accuracy of the Prediction

To compare the received (predicted by ANN) data with the control data (TIM), coefficient of variation (C_v) [10] and Spearman's rank correlation coefficients (r_s) [11] were calculated.

Results

All average predicted values of milk yield for all ANN variants were higher than the control values. Results of the first and fourth ANNs most deviated from the control values, and the average values of the second and third ANNs were the closest to the control ones *(Table 1)*.

A comparison of the results obtained using ANNs showed that individual predictions of cows' milk yield, calculated by ANN-1 were the most precise (deviation from the control value - C_v=5.16%). Least precise were the predictions, calculated by ANN-4 (C_v=7.91). Analysis of rank correlation coefficients, confirm the previous conclusion.

Since the breeding value of cows is usually calculated involving different traits, the predictions for other traits like fat and protein content also were made, because these traits directly affect the price of milk.

Coefficient of variation of differences between control and predicted values of protein content was from 2.17 to 3.07%, for ANN-1, and ANN-4, respectively. Statistically significant Spearman's rank correlation coefficients ($P<0.001$) were established between predicted and control values for all ANNs (r_s ranged from 0.88 to 0.95). ANN-4 was characterized by the smallest r_s (0.888), while the highest value of r_s between predicted and control values had ANN-1 (0.949).

Calculation of prediction of fat content in the milk, showed the advantage in the accuracy of ANN-2 over other variants. The coefficient of variation of ANN-2 was smaller by 1.09, 1.39, and 3.64 percentage points compared to ANN-1, ANN-3, and ANN-4, respectively.

Based on the results of analysis, it was found that trained ANNs are able to predict milk productivity with different accuracy. Some ANNs better predicted milk yield, while others - fat or protein content *(Table 2)*.

Discussion

Since the last 5 years the number of articles dealing with the application of artificial intelligence in animal husbandry has increased significantly [12].

The ANN-1 gave the best prediction for milk yield and protein content. ANN-2 was more suitable for prediction of fat content. ANN-4 showed the worst prediction for all traits, however, this can be explained by the smallest number of MTDMRs that were included in the database for its training (3 MTDMRs from each cow compared to 5 in other ANNs).

An important criterion for farmers is the cost of this assessment. Taking into account the cost of one laboratory analysis of milk quality (0.95 USD), the total cost of analyzing of milk productivity of a single cow during lactation (10 MTDMRs) is 9.5 USD. Using ANNs in

Table 1. *Milk yield, protein, and fat content per standard lactation of cows, calculated by ANN's with different variants of data sets (n=49)*

Parameter	Indicator	TIM	ANN-1	ANN-2	ANN-3	ANN-4
Calculation for Milk Yield	M±Se, kg	8467.1±167.61	8569.1±142.65	8474.8±168.28	8505.7±157.95	8590.4±135.19
	C_v, %	-	5.16	5.98	6.54	7.91
	ΔM, kg	-	102.0	7.7	38.6	123.3
	r_s	-	0.918*	0.886*	0.866*	0.815*
Calculation for Protein Content	M±Se, %	3.24±0.033	3.234±0.028	3.246±0.027	3.261±0.030	3.248±0.029
	C_v, %	-	2.17	2.94	2.83	3.07
	ΔM, %	-	0.0060	0.0056	0.0213	0.0085
	r_s	-	0.949*	0.909*	0.901*	0.888*
Calculation for Fat Content	M±Se, %	4.15±0.061	4.116±0.050	4.139±0.055	4.204±0.051	4.075±0.046
	C_v, %	-	5.40	4.31	5.70	7.95
	ΔM, %	-	0.0345	0.0102	0.0544	0.0751
	r_s	-	0.852*	0.910*	0.853*	0.612*

** Correlation is significant at the 0.001 level; r_s - Spearman's rank correlation coefficients (control: prediction); ΔM - differences between the average values for control and predicted values*

TIM: Test Interval Method; ANN-1: Artificial Neural Network-1; ANN-2: Artificial Neural Network-2; ANN-3: Artificial Neural Network-3; ANN-4: Artificial Neural Network-4

Table 2. Comparative characteristics of accuracy (based on the coefficient of variation) of prediction using different ANN

Variant of Prediction	Traits		
	Milk Yield	Fat Content	Protein Content
ANN-1	1	1	2
ANN-2	2	3	1
ANN-3	3	2	3
ANN-4	4	4	4

1 - the best, 2 - better, 3 - worse, 4 - the worst
ANN-1: Artificial Neural Network-1; ANN-2: Artificial Neural Network-2; ANN-3: Artificial Neural Network-3; ANN-4: Artificial Neural Network-4

various variations, it is possible to reduce the cost of evaluation of one cow by more than 50% (ANN-1, ANN-2, and ANN-3).

The obtained results allow us to conclude that by using ANN it is possible to reduce the number of MTDMRs during the assessment of milk productivity of cows per standard lactation without losing the accuracy of the assessment.

Obviously, the number of MTDMRs, included in the calculations, is one of the important factors affecting the accuracy of the prediction of each ANNs. A similar conclusion was reached by other researchers, because the precision of the models increases with increase in the number of test-day milk records generally [13].

Other important factor is the month of MTDMRs from the beginning of lactation. So, the most correct results were obtained by ANN-1, which used the data of 5 MTDMRs, measured on different periods during lactation: on the first, second, fifth, eighth, and tenth months.

The application of ANN can help to compute (predict) milk productivity of cows for standard lactation if there are missing values in the database [14]. This method is advisable to use in breeding both at the farm level and in breeding centers.

Availability of Data and Materials

Data sets analyzed during the current study are available from the corresponding author (A. Getya) on reasonable request.

Acknowledgments

The authors would like to thank Mr. Oleh Yemets (Dairy management system), Mr. Vitalii Tkach, and Vira Sikorska for their assistance in the data collecting.

Funding Support

This study was conducted as part of the project, supported by the Ministry of education of Ukraine (0119U102345) "Usage of integrated database for breeding merit evaluation of dairy cows" (Researchers A. Getya, M. Matvieiev).

Competing Interests

The authors report no conflicts of interest.

Authors' Contributions

A.G. and M.M. conceived the design of study. M.M and Y.R. collected and analyzed data. M.M., A.G. and Y.R. have approved and read the final version of the manuscript.

References

1. Cole JB, VanRaden PM: Symposium review: Possibilities in an age of genomics: The future of selection indices. *J Dairy Sci*, 101 (4): 3686-3701, 2018. DOI: 10.3168/jds.2017-13335

2. Fessenden B, Weigel DJ, Osterstock J, Galligan DT, Di Croce F: Validation of genomic predictions for a lifetime merit selection index for the US dairy industry. *J Dairy Sci*, 103 (11): 10414-10428, 2020. DOI: 10.3168/jds.2020-18502

3. Miglior F, Fleming A, Malchiodi F, Brito LF, Martin P, Baes CF: A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *J Dairy Sci*, 100 (12): 10251-10271, 2017. DOI: 10.3168/jds.2017-12968

4. Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D: Machine learning in agriculture: A comprehensive updated review. *Sensors (Basel)*, 21 (11):3758, 2021. DOI: 10.3390/s21113758

5. Pollott G, Gootwine E: Appropriate mathematical models for describing the complete lactation of dairy sheep. *Anim Sci*, 71, 197-207, 2000. DOI: 10.1017/S1357729800055028

6. Bentley Instruments I: DairySpec FT User Manual. Bentley Instruments, Inc., Minnesota, USA, 2015.

7. Wiggans G, Shook G: A Lactation measure of somatic-cell count. *J Dairy Sci*, 70 (12): 2666-2672, 1987. DOI: 10.3168/jds.S0022-0302(87)80337-5

8. ICAR: Computing Lactation Yield. Procedure 2 of Section 2 of ICAR Guidelines Computing of Accumulated Lactation Yield. ICAR, 2020.

9. Kingma DP, Ba J: Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)* (arXiv:14126980), 2015.

10. Rawlings JO, Pantula SG, Dickey DA: Applied regression analysis: a research tool. Springer, 1998.

11. de Winter J, Gosling S, Potter J: Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychol Methods*, 21 (3): 273-290, 2016. DOI: 10.1037/met0000079

12. Bao J, Xie Q: Artificial intelligence in animal farming: A systematic literature review. *J Clean Prod*, 331, 2022. DOI: 10.1016/j.jclepro.2021.129956

13. Gorgulu O: Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *South Afr J Anim Sci*, 42 (3): 280-287, 2012. DOI: 10.4314/sajas.v42i3.10

14. Bhosale MD, Singh TP: Comparative study of feed-forward neuro-computing with multiple linear regression model for milk yield prediction in dairy cattle. *Current Science*, 108 (12): 2257-2261, 2015.