

Original scientific paper

APPLICATION OF MACHINE LEARNING IN ESTIMATING MILK YIELD ACCORDING TO THE PHENOTYPIC AND PEDIGREE DATA OF HOLSTEIN-FRIESIAN CATTLE IN SERBIA

LASLO TARJAN^{1*}, IVANA ŠENK¹, DONI PRACNER², LJUBA ŠTRBAC³, MOMČILO ŠARAN³, MIRKO IVKOVIĆ³, NEBOJŠA DEDOVIĆ³

¹University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, Novi Sad, Serbia

²University of Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 3, Novi Sad, Serbia

³University of Novi Sad, Faculty of Agriculture, Trg Dositeja Obradovića 8, Novi Sad, Serbia

*Corresponding author: laci@uns.ac.rs



ISSN 2466-4774
<https://www.contagri.info/>



Submitted: 30.09.2023.
Accepted: 23.11.2023.

SUMMARY

This paper presents a deep neural network (DNN) approach designed to estimate the milk yield of Holstein-Friesian cattle. The DNN comprised stacked dense (fully connected) layers, each hidden layer followed by a dropout layer. Various configurations of the DNN were tested, incorporating 2 and 3 hidden layers containing 8 to 54 neurons. The experiment involved testing the DNN with different activation functions such as the sigmoid, tanh, and rectified linear unit (ReLU). The dropout rates ranging from 0 to 0.3 were employed, with the output layer using a linear activation function. The DNN models were trained using the Adam, SGD, and RMSprop optimizers, with the root mean square error serving as the loss metric. The training dataset comprised information from a unique database containing records of dairy cows in the Republic of Serbia, totaling 3,406 cows. The input parameters (a total of 27) for the DNN included breeding and milk yield data from the cow's mother, as well as the father's ID, whereas the output parameters (a total of 8) consisted of milk yield parameters (a total of 3) and breeding parameters of the cow (a total of 5). Training iterations were conducted using a batch size of 8 over 500, 1000, and 5000 epochs.

Key words:

assessment of milk yield, machine learning, cattle, prediction of breeding values

Abbreviations:

ML – Machine Learning; NN – Neural Network; DNN – Deep Neural Network; ReLU – Rectified Linear Unit; RMSprop – Root Mean Square Propagation; SGD – Stochastic Gradient Descent; DS – data set; ID – Identification Number; MLkg – milk in kilograms; MMpr – percentage of milk fat; PRpr – percentage of milk proteins;

INTRODUCTION

Modern agriculture is currently undergoing substantial transformations to address the challenges posed by global nutrition demands and changing climate conditions. As the world population continues to grow and is expected to exceed 9 billion by 2050 (UN, 2019), agricultural sectors are under pressure to enhance food production while minimizing resource usage (Foley et al., 2011). In light of these challenges, animal husbandry, as an integral component of the agricultural system, must evolve towards adopting more efficient and sustainable production methods. The introduction of machine learning into agricultural practices marks a significant leap towards modernization and the enhancement of efficiency in this sector (Chafai et al., 2023; Nayeri et al., 2019).

Milk has evolved into a fundamental source of protein, vitamins, and minerals for people worldwide (Walstra et al., 2005). Consequently, the efficiency of milk production and the quality of dairy products have a direct impact on global nutrition (FAO, 2018). Enhancing the genetic basis of dairy cows can result in higher milk yields, improved milk composition, and increased animal resistance to diseases (Dekkers, 2004). In Serbia, as in numerous other countries, the ongoing initiatives persist to enhance the genetic potential of dairy cows through selective breeding and genetic evaluations (Stanojević et al., 2015). The adoption of modern technologies, such as genomic selection, facilitates accelerated progress and more precise predictions of the future performance of animals (Hayes et al., 2009).

Selection programs establish criteria for parental selection based on statistical analysis of extensive databases, utilizing compiled data on ancestry and phenotypic measurements to assess the breeding values of potential candidates (Štrbac et al., 2023). In recent years, traditional selection methods, grounded in quantitative genetics, have increasingly incorporated molecular genetics analyses (Štrbac et al., 2023). In addition to conventional statistical evaluation methods, the adoption of machine learning (ML) techniques has gained prominence, facilitating the development of rapid programs and unveiling previously undiscovered relationships (Shahinfar et al., 2012; Tarjan et al., 2021).

The Holstein-Friesian breed is the predominant dairy breed within the registered cattle population in Serbia, with the majority located in the AP Vojvodina. According to data from the Central Breeding Organization, a total of 67,671 head of the Holstein-Friesian cattle were registered in this area in 2021 (Štrbac et al., 2023). The average milk production during a 305-day lactation period in the same year amounted to 7,405 kg per cow (Štrbac et al., 2023). A comparison of this figure with the milk yield of cows documented in other ICAR member countries (ICAR, 2023) reveals potential opportunities for enhancing the traits of dairy population in Serbia. Improving cow selection methods has been recognized as one of the viable pathways toward achieving this objective (Štrbac et al., 2023).

The primary objective of this research is to demonstrate the potential application of machine learning (ML) in estimating the milk yield of Holstein-Friesian cows in Serbia according to the phenotype and pedigree of their parents. Notably, this breed represents a population with a relatively small number of individuals, which can be a limiting factor when training neural networks (NN). To address this challenge, a Deep Neural Network (DNN) was employed, comprising dense layers with various configurations and activation functions. The DNN was trained using the Adam, SGD, and RMSprop optimizers, employing the root mean square error as the loss metric. The training dataset consisted of information obtained from a unique database containing records of a total of 3,406 dairy cows in the Republic of Serbia.

MATERIAL AND METHODS

To explore the viability of utilizing machine learning (ML) in estimating the milk yield of future cow generations based on the parental phenotype and pedigree data, the initial step involved selecting a dataset (DS) containing all relevant information (including parental phenotypic and milk yield data). Subsequently, the dataset was divided, designating a portion for training the Deep Neural Network (DNN) and another for testing purposes. Following this, configurations for the DNN were established, and the training process was executed. Finally, the trained DNN underwent testing using a section of the DS exclusively reserved for testing, ensuring it had not been utilized in the training phase of the DNN.

Data Sets and Quality Control

The dataset employed for training and testing the DNN was generated using the web application BioITGenoSelect (<https://bioitgenoselect.polj.uns.ac.rs/genoselect>, accessed on February 10, 2022) for the specified animals. This dataset was created based on imported information sourced from databases managed by the Faculty of Agriculture within the Central Breeding Organization in the livestock sector of the Vojvodina Province, Serbia. Data were specifically selected for animals with fully available pertaining information on their parents and the phenotype and

first lactation milk yield data for both the individual animal and its parents. Following these selection criteria, a dataset (DS) was formulated, comprising information from 3,406 animals, which was subsequently utilized for the training and testing procedures of the DNN (the dataset is accessible at: <https://oblak.uns.ac.rs/index.php/s/R46gQFj8Dt6A4aa>). Alongside the phenotypic and milk yield data, relative reproductive parameters derived from processing phenotypic information via the WOMBAT software (Meyer, 2007) were also integrated. These parameters served as an evaluation metric for the trained DNN. Prior to deployment into the Neural Network, all the data within the DS were normalized within the range of 0 to 1. The DS was partitioned into two segments: 80% allocated for training purposes and the remaining 20% reserved for testing.

Deep Neural Network

The DNN was utilized for milk yield prediction, employing dense (fully connected) layers arranged sequentially with dropout layers incorporated after each hidden layer. Configurations involving 2 and 3 hidden layers of 8 to 54 neurons were selected. The DNN utilized a total of 27 input parameters, consisting of reproductive parameters obtained from WOMBAT, the mother's milk yield parameters, and the identification number (ID) of the father (refer to Fig. 1 for the input file layout). Additionally, there were 8 output parameters, including milk yield parameters (3 in total: MLkg, MMpr, PRpr) and the breeding parameters of the head of cattle obtained from WOMBAT (refer to Fig. 2 for the output file layout). Various activation functions such as the sigmoid, tanh, and Rectified Linear Unit (ReLU) were employed. Dropout rates were set within the range of 0 to 0.3, and a linear activation function was used for the output layer.

Training and testing

The Adam, SGD, and RMSprop optimizers were employed, and the root mean square error served as the loss metric during DNN training. The model underwent training with a batch size of 8 across 500, 1000, and 5000 epochs. The training and testing procedures were conducted using Python, utilizing the Keras open library for the DNN, in conjunction with the TensorFlow open library for machine learning (Chollet, 2017). After each DNN training cycle, testing was conducted. The ensuing results are detailed and analyzed in the subsequent chapter. The output DNN models were stored for potential reuse or future integration.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	otac_i	majka-uot	majka-uoo	majka-vk	majka-ll	majka-sg	majka-dt	majka-pkar	majka-sk	majka-uom	majka-ug	majka-uon	majka-pznon	majka-pznss	majka-up	majka-uov	majka-vpv	majka-pps	majka-dps	majka-dv	majka-vzvz	majka-jcl	majka-pzs	majka-dzs	majka-MLkg	majka-MMpr	majka-PRpr
2	13	80	83	8	7	6	7	7	6	79	7	80	7	6	5	79	7	5	5	6	5	5	5	4	7772	3.43	3.09
3	13	78	81	7	8	5	6	4	7	74	6	81	7	5	4	76	7	5	7	7	6	7	6	6	8132	2.85	2.87
4	13	85	84	8	7	6	7	7	7	79	7	87	8	5	6	86	8	5	5	5	8	8	5	5	5645	3.35	3.15
5	13	80	79	6	6	6	6	7	7	79	7	79	7	5	4	81	6	5	5	6	8	8	7	5	8315	3.54	3.21
6	13	77	80	7	6	6	5	3	7	79	7	75	7	6	3	77	7	4	6	7	5	4	5	5	8507	3.83	3.22
7	13	75	79	7	6	6	5	3	6	79	7	72	6	7	4	74	5	5	7	8	6	5	6	5	8438	3.4	3.34
8	13	77	85	8	7	6	5	5	7	70	5	70	6	3	3	78	7	5	6	7	5	5	5	5	8289	3.86	3.52
9	13	81	84	7	7	7	7	7	7	84	8	79	8	6	4	79	7	5	5	6	7	8	8	4	6817	3.8	3.14
10	13	80	82	8	7	6	7	7	7	84	8	81	6	5	5	78	7	5	5	7	7	6	8	5	5972	4.03	3.02
11	13	76	76	7	5	5	5	4	7	79	7	81	7	5	4	73	7	4	6	7	4	5	4	7	8558	3.22	3.12
12	13	78	79	7	6	5	5	4	6	79	7	77	8	6	3	78	7	5	6	7	6	5	5	4	7373	3.69	2.87
13	13	77	77	8	6	5	4	7	6	74	6	81	7	5	4	75	7	4	6	7	6	5	5	4	7803	3.36	3.41
14	13	77	80	6	6	6	6	6	6	65	4	74	7	3	4	79	7	4	7	5	7	7	5	7	7998	3.19	3.11
15	13	78	76	6	5	5	6	7	7	84	8	76	7	6	4	79	7	6	5	6	8	8	9	4	8243	3.7	3.28

Figure 1. An example of input data for the DNN provided in the form of a *.xlsx file

	A	B	C	D	E	F	G	H
1	MLkg	MMpr	PRpr	uot	uoo	uom	uon	uov
2	8530.565	3.531081	3.260936	79.87325	81.66663	79.14051	79.22282	79.26806
3	8798.247	3.551077	3.244011	77.8919	80.89771	76.93843	76.25176	77.22449
4	6142.244	3.737818	3.343989	82.24869	84.06188	81.55835	80.78367	82.42427
5	8264.982	3.649944	3.20855	79.1768	81.73261	79.32081	78.0676	78.25447
6	8524.694	3.873342	3.240335	77.76213	80.78734	76.45969	75.80959	77.27591
7	8188.531	3.730812	3.240314	78.3194	80.75447	77.5975	76.79392	78.03668
8	7812.046	3.651052	3.328523	77.11378	80.80462	75.97302	75.30981	76.13248
9	7910.359	3.784709	3.206102	79.3097	81.48425	78.42081	78.09421	78.88959
10	7500.398	3.903578	3.256206	79.15217	81.53278	78.79627	77.50158	78.65194
11	8892.08	3.493149	3.245575	77.24812	80.84675	75.67873	75.31637	76.43225
12	8184.33	3.841308	3.29128	76.42672	79.3715	72.55666	74.8471	76.34967
13	7858.496	3.536352	3.333364	76.50883	80.19426	75.42416	74.69494	75.37053
14	8652.293	3.716739	3.247727	78.33858	81.14575	78.64311	77.35667	77.043
15	8457.733	3.07052	3.093247	78.68154	80.75445	83.75188	76.45671	77.314

Figure 2. An example of the output provided in the form of a csv file

RESULTS AND DISCUSSION

After completing the DNN training phase, the model was saved to a *.keras file for future use. The trained model was then assessed using a portion of the dataset that had been set aside for testing purposes. Various optimizers such as the Adam, SGD, and RMSprop, accompanied by activation functions like the sigmoid, tanh, and ReLU, were employed during the DNN training across 500, 1000, and 5000 epochs. The DNN training encompassed configurations with 2 and 3 hidden layers, and the resulting predictions were organized into a table. Through a comprehensive search, the most optimal hyperparameters were identified for the model utilizing 2 hidden layers with a neuron configuration of - 54 - 8 - and implementing a dropout rate of 0.1.

Table 1 presents the maximum absolute prediction error, displayed as a percentage, for each DNN. Additional results can be accessed in the table via the following link: <https://oblak.uns.ac.rs/index.php/s/R46gQFj8Dt6A4aa>. In scenarios where a single model is utilized for all three observed parameters (MLkg, MMpr, PRpr) as depicted in Table 1, the best outcomes from both variants of the DNN (with 2 or 3 layers) are highlighted in bold font. Conversely, if distinct models were employed for each observed parameter, the superior results achieved would be indicated using blue, yellow, and green colors. Using the results obtained, the two most promising variants of the DNN model were chosen. Subsequently, predictions for the milk yield parameters of the offspring were made for 100 randomly selected pairs of parents from the test subset of the dataset. The outcomes of these predictions are presented in Figures 3 and 4.

When considering the scenario where all parameters are determined using a single DNN model, it is evident from Table 1 that the most optimal results are achieved by the DNN featuring 2 hidden layers. This model utilizes a combination of the RMSprop optimizer with a sigmoid activation function, trained over 500 epochs. The identified model yields a prediction accuracy of 76.1% for milk yield (as depicted in Fig. 3). This accuracy aligns with the findings in Table 1, highlighting the largest error of 23.9% for the MLkg parameter in the aforementioned model.

Table 1. Maximum absolute prediction error for the DNN

Opt	Act	epochs output	3 hidden layers			2 hidden layers		
			500	1k	5k	500	1k	5k
Adam	sigmoid	MLkg	35.33%	25.17%	35.25%	23.78%	30.02%	38.72%
		MMpr	17.36%	17.88%	28.86%	17.57%	16.78%	19.66%
		PRpr	10.83%	10.60%	8.42%	6.22%	25.94%	10.98%
	tanh	MLkg	41.75%	29.97%	32.91%	25.37%	31.75%	31.62%
		MMpr	27.75%	15.05%	17.72%	18.48%	17.84%	17.81%
		PRpr	31.28%	6.58%	26.25%	6.33%	25.64%	10.40%
	relu	MLkg	33.28%	35.07%	41.54%	27.46%	29.43%	31.44%
		MMpr	28.75%	18.80%	17.98%	16.38%	16.04%	18.14%
		PRpr	7.27%	6.70%	31.71%	6.50%	25.49%	11.18%
SGD	sigmoid	MLkg	42.25%	30.70%	30.38%	28.85%	30.37%	30.57%
		MMpr	17.85%	19.74%	17.64%	18.75%	22.29%	19.47%
		PRpr	32.26%	7.34%	7.72%	6.35%	24.98%	10.88%
	tanh	MLkg	28.04%	35.92%	27.99%	25.24%	30.70%	30.03%
		MMpr	18.36%	19.02%	17.84%	17.05%	18.50%	18.03%
		PRpr	7.88%	7.18%	25.47%	6.83%	25.58%	10.67%
	relu	MLkg	36.93%	31.97%	31.40%	70.74%	70.74%	31.10%
		MMpr	21.10%	19.91%	18.15%	17.29%	19.15%	17.94%
		PRpr	25.27%	8.01%	6.88%	7.04%	26.00%	11.06%
RMSprop	sigmoid	MLkg	33.68%	34.39%	39.25%	23.93%	29.83%	33.23%
		MMpr	15.43%	16.83%	15.91%	17.01%	16.14%	17.93%
		PRpr	7.50%	25.01%	10.88%	6.32%	24.91%	10.76%
	tanh	MLkg	31.95%	39.39%	35.33%	26.29%	29.72%	30.00%
		MMpr	26.51%	17.84%	17.53%	18.15%	19.12%	17.24%
		PRpr	25.34%	32.71%	24.78%	6.30%	26.08%	9.66%
	relu	MLkg	26.83%	34.93%	26.16%	27.04%	31.98%	30.73%
		MMpr	18.34%	15.96%	19.12%	16.11%	18.70%	16.90%
		PRpr	24.94%	25.35%	7.94%	6.15%	25.48%	10.75%

When employing multiple models for each specific output parameter, the following optimal configurations should be considered:

- For the parameter MLkg, the recommended combination involves using the Adam's optimizer with a sigmoid activation function trained over 500 epochs.
- When predicting MMpr, the preferable model configuration comprises the utilization of RMSprop optimizer with a ReLU activation function across 500 epochs.
- Regarding the PRpr parameter, the most suitable model setup involves employing the RMSprop optimizer alongside a sigmoid activation function, trained over 500 epochs.

When employing a combination of several models, an accuracy of 76.2% (Fig. 4) is achieved, representing only a marginal 0.1% improvement compared to the previous case. Despite this slight enhancement, it is advisable to opt for utilizing separate models for each parameter. This approach is recommended to attain more precise predictions, crucial for making the best selection of parent pairs based on accurate predictions.

Using the best DNN models, user applications can be developed to forecast the milk and breeding parameters of future calves based on parental input data. Notably, the output files should present absolute values of the parameters instead of normalized ones. This adjustment enables easy comparison with the input data. Similar treatment of output data was applied in the example of the *.csv file (Fig. 2).

The previous analysis focused on the results derived from the maximum absolute error. However, if we consider the maximum mean absolute error, the findings are as follows:

- For the 2-layer DNN trained with the SGD optimizer, activated with the ReLU function, and trained over 5000 epochs, the maximum mean absolute error is 6.09%.
- In the case of the 3-layer DNN trained with the RMSprop optimizer, activated with the sigmoid function, and trained over 500 epochs, the maximum mean absolute error is 5.83%.

These error percentages, while relatively high, are acceptable considering the size of the dataset used. However, it is important to note that these values exceed the desired thresholds for model performance. Ideally, for a well-performing neural network, we would anticipate an absolute maximum error of less than 5% and a mean absolute error below 1%. Evaluating the prediction errors made across the training dataset reveals slightly better results, but they also fall short of the expected ranges. This suggests the necessity for a larger training dataset to enhance the performance of the DNN.

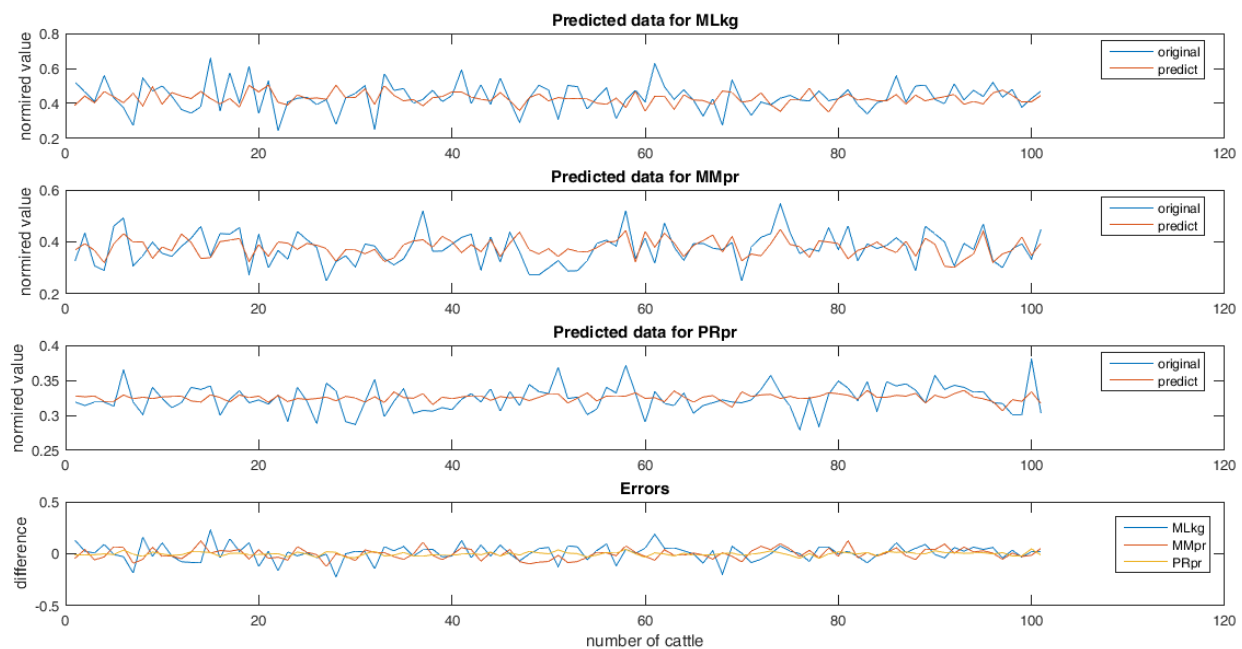


Figure 3. Plot of the milk yield parameters and associated errors for 100 randomly selected pairs of parents predicted using the RMSprop sigmoid 500 2-layer model.

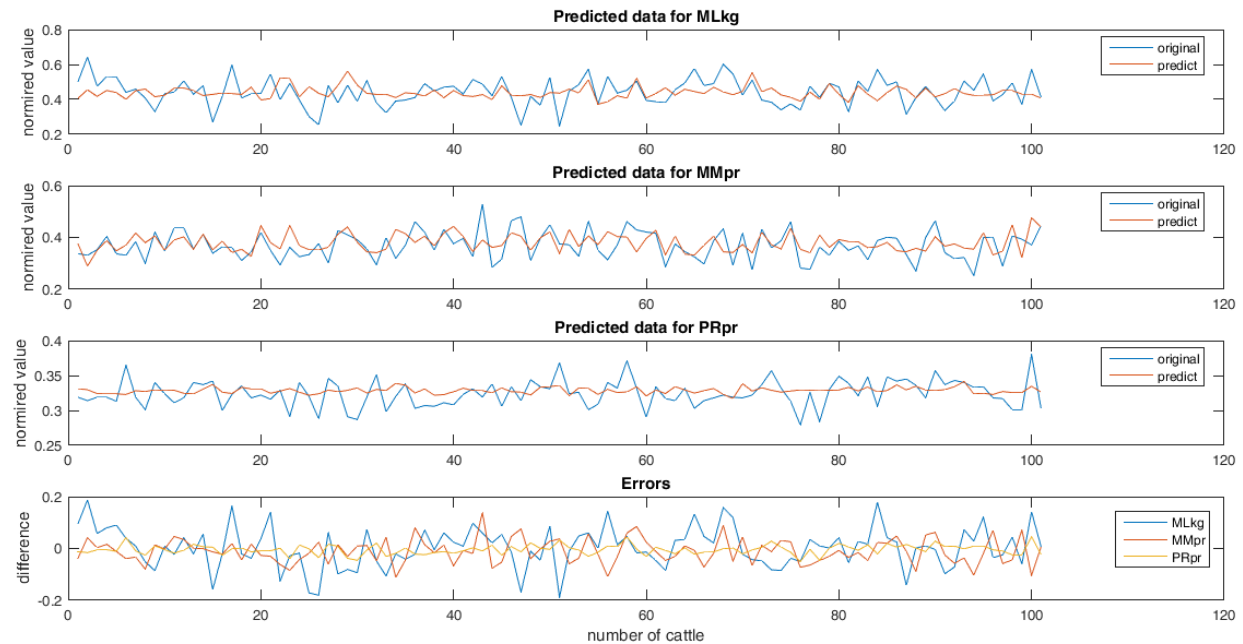


Figure 4. Plot of the milk yield parameters and associated errors for 100 randomly selected pairs of parents predicted using the combined 2-layer model.

CONCLUSION

This paper showcases the application of artificial intelligence in estimating the milk yield of future cow generations through machine learning techniques. The study utilized the data from a unique database encompassing dairy cows in the Republic of Serbia for training the Deep Neural Network (DNN) (a dataset comprising 3,406 cows at the current juncture). Considering the theoretical framework advocating that deep learning necessitates a dataset size at least 10 times larger than the number of variables (González-Recio et al., 2014; Morota et al., 2018), the observed limitations in the testing results of the trained DNN suggest that this number of cows might be insufficient.

Among the analyzed configurations of the DNN, the highest achieved accuracy in predicting the milk yield of the offspring was 76.2%, which falls approximately twenty percent below the anticipated level. Nevertheless, this level of accuracy remains acceptable, especially considering the context of being in a small country with a limited population. These neural networks can indeed serve in future endeavors, aiding in data generation that benefits farmers. Additionally, they can assist in verifying the accuracy of data input for future livestock.

To enhance the performance of the model, it is essential to incorporate data from a larger pool of individuals. This expansion can be achieved by broadening the dataset to encompass information on livestock from the geographic region to which Serbia belongs. By including the data from this larger population, a more substantial number of individuals can be integrated into the training phase of the DNN.

Acknowledgements: The data used in this research were collected during the implementation of the BioITGenoSelect project (No: 6066512) within the PROMIS program funded by the Science Fund of the Republic of Serbia and this paper was supported by the Ministry of Science, Technological Development and Innovation under Project No. 451-03-47/2023-01/200156 entitled “Innovative scientific and artistic research from the FTS (activity) domain”.

Conflict of interest: The authors declare that they have no conflict of interest.

REFERENCES

- Chafai N., Hayah I., Houaga I., Badaoui B. (2023): A review of machine learning models applied to genomic prediction in animal breeding. *Frontiers in Genetics*, 14: 1150596. <https://doi.org/10.3389/fgene.2023.1150596>
- Chollet F. (2017): *Deep Learning with Python*. Manning Publications Co.
- Dekkers J.C.M. (2004): Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science*, 82 E-Suppl: E313-328. https://doi.org/10.2527/2004.8213_supplE313x
- FAO (2018): The state of Food and Agriculture. In: *The State of the World*. <https://www.fao.org/3/i9549en/19549EN.pdf>

- Foley J.A., Ramankutty N., Brauman K.A., Cassidy E.S., Gerber J.S., Johnston M., Mueller N.D., O'Connell C., Ray D.K., West P.C., Balzer C., Bennett E.M., Carpenter S.R., Hill J., Monfreda C., Polasky S., Rockström J., Sheehan J., Siebert S., Tilman D., Zaks D.P.M. (2011): Solutions for a cultivated planet. *Nature*, 478(7369): 337-342. <https://doi.org/10.1038/nature10452>
- González-Reco O., Rosa G.J.M., Gianola D. (2014): Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livestock Science*, 166: 217-231. <https://doi.org/10.1016/j.livsci.2014.05.036>
- Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009): Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2): 433-443. <https://doi.org/10.3168/jds.2008-1646>
- ICAR (2023): The Global Standard for Livestock Data. Statistics 2023. Available at: <https://my.icar.org/stats/list> (accessed on 17 January 2023).
- Meyer K. (2007): WOMBAT - A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University SCIENCE B*, 8(11): 815-821. <https://doi.org/10.1631/jzus.2007.B0815>
- Morota G., Ventura R.V., Silva F.F., Koyama M., Fernando S.C. (2018): Big Data Analytics and Precision Animal Agriculture Symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture 1. *Journal of Animal Science*, 96(4): 1540-1550. <https://doi.org/10.1093/jas/sky014>
- Nayeri S., Sargolzaei M., Tulpan D. (2019): A review of traditional and machine learning methods applied to animal breeding. *Animal Health Research Reviews*, 20(1): 31-46. <https://doi.org/10.1017/S1466252319000148>
- Shahinfar S., Mehrabani-Yeganeh H., Lucas C., Kalhor A., Kazemian M., Weigel K.A. (2012): Prediction of Breeding Values for Dairy Cattle Using Artificial Neural Networks and Neuro-Fuzzy Systems. *Computational and Mathematical Methods in Medicine*, 127130. <https://doi.org/10.1155/2012/127130>
- Stanojević D., Đedović R., Bogdanović V., Popovac M., Perišić P., Beskorovajni R., Lazarević M. (2015): The potentials of using selection index in the assessment of breeding values of Holstein breeds in Serbia. *Biotechnology in Animal Husbandry*, 31(4): 523-532. <https://doi.org/10.2298/bah1504523s>
- Štrbac L., Pracner D., Šaran M., Janković D., Trivunović S., Ivković M., Tarjan L., Dedović N. (2023): Mathematical Modeling and Software Tools for Breeding Value Estimation Based on Phenotypic, Pedigree and Genomic Information of Holstein Friesian Cattle in Serbia. *Animals*, 13(4): 597. <https://doi.org/10.3390/ani13040597>
- Tarjan L., Šenk I., Pracner D., Rajković D., Štrbac L. (2021): Possibilities for applying machine learning in dairy cattle breeding. 20th International Symposium INFOTEH - Jahorina (INFOTEH), 1-6. <https://doi.org/10.1109/INFOTEH51037.2021.9400672>
- UN (2019): World Population Prospects 2019. Available at: https://reliefweb.int/attachments/f46d3fce-97bb-327f-a065-a813f9969af7/WPP2019_Highlights.pdf
- Walstra P., Walstra P., Wouters J.T.M., Geurts T.J. (2005): *Dairy Science and Technology* (2nd ed.). CRC Press, Taylor & Francis. <https://doi.org/10.1201/9781420028010>