Turkish Journal of Agricultural Engineering Research

https://dergipark.org.tr/en/pub/turkager
https://doi.org/10.46592/turkager.2020.v01i02.011

Turk J Agr Eng Res (TURKAGER)
e-ISSN: 2717-8420
2020, 1(2): 354-367

Research Article

Evaluation of Normalization Techniques on Neural Networks for the Prediction of 305-Day Milk Yield

Asli AKILLI[ID][a*] **Hulya ATIL**[ID][b]

[a]*Department of Animal Science, Faculty of Agriculture, University of Kirsehir Ahi Evran, 40200, Bagbası- Kirsehir-TURKEY*

[b]*Department of Animal Science, Faculty of Agriculture, University of Ege, 35040, Bornova- Izmir- TURKEY*

(*): Corresponding author, asliakilli@ahievran.edu.tr; Tel: +90-386-2804841, Fax: +90-386-2804832

ABSTRACT

In this study, the impact of data preprocessing on the prediction of 305-day milk yield using neural networks were investigated with regard to the effect of different normalization techniques. Eight normalization techniques "Z-Score, Min-Max, D-Min-Max, Median, Sigmoid, Decimal Scaling, Median and MAD, Tanh-Estimators" and five different back propagation algorithms "Levenberg-Marquardt (LM), Bayesian Regularization (BR), Scaled Conjugate Gradient (SCG), Conjugate Gradient Back propagation with Powell-Beale Restarts (CGB) and Brayde Fletcher Gold Farlo Shanno Quasi Newton Back propagation (BFG)" were examined and tested comparatively for the analysis. Neural network architecture was optimized and tested with several experiments. Results of the analysis show that applying different normalization techniques affect the performance and the distribution of outputs influences the learning process of the neural network. The magnitude of the effects varied with the type of back propagation algorithms, activation functions, and network's architectural structure. According to the results of the analysis, the most successful performance value in the 305-day milk yield estimation was obtained by using the neural network structured by using the Decimal Scaling normalization technique with the Bayesian Regulation algorithm (R^2_{Adj} = 0.8181, RMSE= 0.0068, MAPE= 160.42 for test set; R^2_{Adj} =0.8141, RMSE= 0.0067, MAPE= 114.12 for validation set).

RESEARCH ARTICLE

Received: 20.07.2020
Accepted: 07.09.2020

Keywords :

- 305-day milk yield,
- Agricultural data,
- Back propagation algorithms,
- Data pre-processing,
- Neural network,
- Normalization

To cite: Akilli A, Atil H (2020). Evaluation of Normalization Techniques on Neural Networks for the Prediction of 305-Day Milk Yield. Turkish Journal of Agricultural Engineering Research (TURKAGER), 1(2): 354-367. *https://doi.org/10.46592/turkager.2020.v01i02.011*

INTRODUCTION

One of the most important objectives of the studies carried out in livestock science is to reduce the costs of breeding and feeding by optimizing and to make evaluations in this direction. For this purpose, milk yield predictions are made using various statistical and artificial intelligence-based methods. In evaluations on dairy cattle, analysis using test day records provides important advantages on the prediction of 305-day milk yield (Mostert *et al.,* 2006; Dongre *et al.,* 2012). The use of test day records is very useful for dairy cattle breeders in decreasing cost of milk recording, making genetic evaluations about animals in early stages of lactation, culling unproductive cows and improving production pattern (Dongre *et al.,* 2012). Estimates of milk yield on herd or individual basis strongly affect energy consumption based on daily milk production, plant utilization and farm income, and provide large benefits to dairy industry at farm level. Accurate and effective usage of milk yield estimation methods can contribute to the correct modeling of milk production patterns at certain time periods and to adapt to factors affecting the supply-demand balance on a farm basis (Murphy *et al.,* 2014). However, the 305-day milk yield estimate provides a basis for the genetic evaluation of dairy cattle (Grzesiak *et al.,* 2003). In the production of high-yielding dairy cows, the selection of genetically superior bulls is very important and milk yield prediction models are used in this process. Early detection of superior bulls accelerates the production process, which continues with semen collection and insemination, so that gains in genetic progress can be achieved (Sharma *et al.,* 2007). Conventional regression models have been widely used as prediction tools for 305-day milk yield. Today, with the technological advance on computer systems, neural network method can be used as an alternative to statistical analysis methods and they are the subject of many successful studies. Neural network method is designed by as an example of the working structure of the human brain and the learning function is performed by using experiences similar to humans. Neural networks provide solutions to new problems faced in the future through generalization capability.

Neural networks method is one of the most popular artificial intelligence methods which are widely used in animal husbandry as well as in many applied sciences (Akıllı, 2019). Neural networks have been used for milk yield estimation studies in the field of dairy science (Salehi *et al.,* 1998; Sanzogni and Kerr, 2001; Grzesiak *et al.,* 2003; Grzesiak 2006; Sharma *et al.,* 2006; Sharma *et al.,* 2007; Hosseinia *et al.*, 2007; Edriss *et al.,* 2008; Njubi *et al.,* 2009; Gandhi *et al.,* 2010; Njubi *et al.,* 2010; Ruhil *et al.,* 2011; Dongre *et al.,* 2012; Gandhi *et al.,* 2012; Tahmoorrespur *et al.,* 2012; Murphy *et al.,* 2014; Kong *et al.,* 2018) and ruminant animals (Ominakis *et al.,* 2002; Torres *et al.,* 2005; Ince and Sofu, 2013, Karadas *et al.,* 2017). In the 305-day milk yield estimation examined in the context of linear regression analysis, information on reproductive activities such as calving interval as well as milk yield on test day and lactation information are included in the model structures (Grzesiak *et al.,* 2006; Sharma *et al.,* 2007; Dongre *et al.,* 2012; Takma *et al.,* 2012; Görgülü, 2012).

Environmental and genetic factors are highly effective on the milk yield data pattern. Various limitations can be encountered in the measurement process of the data. These limitations, which directly affect the distribution of the data structure, can cause noise and inconsistencies. At this point, the data can be resized using normalization methods. The use of normalization methods can provide considerable benefits in terms of detecting anomalies in the data distribution and using the model under study more effectively and efficiently. Data preprocessing and normalization techniques provide significant improvements in neural network performance. In this context, normalization techniques contribute to the transformation of neural network inputs according to the data range (Logistics, Tanh-Sigmoid) of the defined activation function.

In neural network analysis, the normalization process is of great importance in data structures where the number of observations and variables is large, especially where the dimensional differences in input observation values are included. The impact of normalization techniques on neural network performance, their characteristics, and learning processes have been discussed. According to neural network literature, normalization can be useful for learning process, and it may be essential, to enable them to detect patterns contained in the learning data set. Normalization of inputs or rescaling data could sometimes greatly help to use prior knowledge and to reduce the complexity of the data structure (Lacroix *et al.,* 1997). Some studies have shown that different normalization methods have been affected the performance of the neural network (Shanker *et al.,* 1996; Jain *et al.,* 2005; Jayalakshmi and Santhakumaran, 2011; Panigrahi and Behera, 2013).

In this study, the impact of eight normalization techniques "Z-Score, Min-Max, D-Min-Max, Median, Sigmoid, Decimal Scaling, Median and MAD, Tanh-Estimators" on neural network performance were examined comparatively for the milk yield prediction. In this context, it was aimed that determine the impact of distributing the input vectors on an equal basis with respect to the output pattern on the learning of neural networks.

MATERIALS and METHODS

Data Source

Data sets contain milk yield information on Turkey Cattle Breeders Central Association registered Holstein Friesian cows. The Study material consists of 1000 data records. In this study, neural networks were studied with different normalization techniques in order to estimate the 305-day milk yield. Test day data, age of first calving, lactation number, and days in milk (DIM) were used, in order to analyze the impact of the distribution of the outputs when predicting 305-day milk yield. Input variables of analysis were determined as the first four test day record, age of the first calving, lactation number, and DIM in designed system. The output of the system is defined as 305-day milk yield. Table 1 shows the descriptive statistics of the variables in the data set. Accordingly, skewness and kurtosis values indicate that the dependent variable does not show normal distribution. After logarithmic transformation, 305-day milk yield values were normally distributed. When the descriptive statistics of the other variables in data set are examined, it is seen that there are no outliers. Neural network analyses were examined on different parameter combinations to determine the optimal values of the model parameters. Hundreds of parameter combinations setting performed for neural network at each phase of training, testing and validating for data set. All experiments were done using MATLAB (R2016a).

Table 1. Descriptive statistics of data set

Variables	Mean	Std. Dev.	CV[1]	Min	Max	Skewness	Kurtosis
305-Day Milk Yield	6173.37	1605.11	26.0006	2033	13432	0.740	1.339
Calving Interval	74.365	23.9587	32.2177	23	101	-0.759	-0.841
Lactation Number	4.497	1.7975	39.9724	1	6	-0.765	-0.92
DIM[2]	284.052	18.143	6.38721	195	305	-1.763	3.986
TestDay1	25.1801	6.9058	27.4257	8	56.9	0.556	0.445
TestDay2	24.9515	7.0633	28.3081	6	54.5	0.646	0.977
TestDay3	23.8841	6.9252	28.9950	5	58.7	0.839	2.015
TestDay4	22.4664	6.7920	30.2319	4	54	0.828	2.097

Number of Data Records: 1000; [1]CV: Coefficient of Variation (%). [2]DIM: Days in Milk

Multilayer Perceptron

Machine learning studies, which are an important part of artificial intelligence, give computers the behavior of working with experimental data. Neural networks method is one of the methods studied in the context of machine learning. The method, which has been used successfully by researchers in many different fields, was developed on the basis of the working principles of neurons in the human brain (Zhang *et al.,* 1998). Neural networks collect information from the environment during the training process and store this information through various connections, like the human brain. The learning function of neural networks takes place in a similar way to human experience. This method enables the solution of new problems in the future by means of the information learned. The first studies on neural networks started in 1943 with the theory of Warren McCulloch and Walter Pitts about the functioning of neurons. Nowadays, the studies on the theory and application of neural networks are increasing and showing very successful results. The advances in technological developments and the fact that computers have an increasing graph in technical terms have a positive effect on the development of neural networks (Akıllı, 2019).

In this study, multilayer perceptron (MLP) neural network with back propagation of error learning mechanism was used for the neural network architecture. MLP utilizes a supervised learning strategy (Negnevitsky, 2002; Panigrahi and Behera, 2013). The back propagation learning algorithms can be used for neural network training. Basically, propagation and update steps have been used for the back propagation algorithms (Savegnago *et al.,* 2011). The back propagation process is repeated until the error criterion (sum of error squares) reaches the specified level. In back propagation algorithm feed forward process, w_{kj} is the weight value of the link that connects the k^{th} input layer process element to the j^{th} hidden layer element. Where, o_k is the output of k^{th} process element in the input layer and w_{kj} is the synaptic weight. Net input (Net) is given in Equation 1.

$$Net_j^a = \sum_{k=1}^{n} w_{kj} o_k \tag{1}$$

$$o_j = \frac{1}{1 + e^{-(Net_j^a + \beta_j^a)}} \tag{2}$$

The Net input obtained is passed through the activation function so that the output of the neuron in the hidden layer is calculated. In this study, sigmoid activation function is discussed. The output is expressed as in Equation 2. Where, o_j is the output of j^{th}

hidden layer element and β_j is refers to the weight of the threshold element that connects to the j^{th} hidden layer element (Öztemel, 2002).

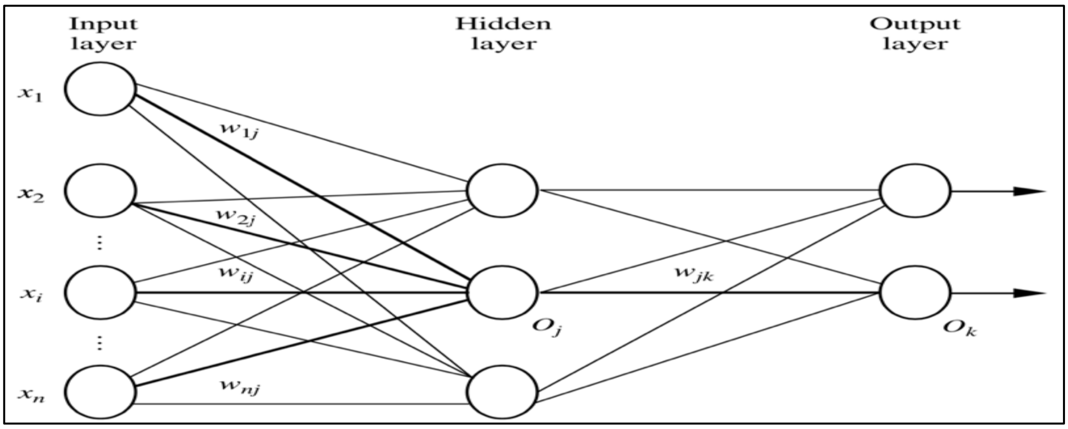

Figure 1. Multilayer perceptron (Han and Kamber, 2006)

Figure 1 shows multilayer perceptron. In the present study, neural network was trained and simulated using five different back propagation algorithms Levenberg-Marquardt (LM), Bayesian Regularization (BR), Scaled Conjugate Gradient (SCG), Conjugate Gradient Backpropagation with Powell-Beale Restarts (CGB) and Brayde Fletcher Gold Farlo Shanno Quasi Newton Backpropagation (BFG) up to 10000 epochs or till the algorithms are truly converged. Also, two different activation functions (Tan-Sig, Log-Sig) were used to compute the output from the summation of weighted inputs of neurons in each hidden layer (Dongre *et al.,* 2012). Network parameters such as learning rate (0.01), momentum (0.05-0.95) were used with different combinations. The network was tested with the different numbers of hidden layer (1-3) and neuron (3-20). Initial weights and bias matrix were randomly determined. In the present study, neural network was optimized by training and testing errors in order to investigate the overfitting and underfitting problems. Training set was determined randomly with 80% for this case. The accuracy of normalization techniques was measured at each phase of modeling. Hundreds of combinations of neural network parameter setting have been designed for prediction models at each phase of training, testing and validation for data set.

Normalization Techniques

In present study, it was aimed to investigate the importance of data normalization, especially when applying the neural network approach to 305-day milk yield prediction problem. Thus, the distribution of the output pattern in the training data set was analyzed with regard to its effect on learning. Data normalization provides a comparable range for input and output pattern in neural network analysis and significantly affects equally distributing the importance of each input variable in the analysis. Thus, distributing the input vectors uniformly with respect to the output pattern of all values was provided in the training of the neural network. In the normalization process, data is scaled in specific ranges such as "-1.0- 1.0", "0.0- 1.0" or "0.1- 0.9" through various techniques. In the present study, all normalization techniques were applied to all the feature vectors for given data, and then neural network was trained with the created training set.

In Z-Score normalization technique, data values are normalized according to their mean and standard deviation. The most important advantage of this normalization technique is that it can be used effectively to reduce the effect of outliers in the data set. However, this method may be useful when the actual minimum and maximum value of variables are unknown. But it does not perform well with non-stationary time series due to the variation of the standard deviation and mean depending on the data structure and time (Han and Kamber, 2006; Shalabi *et al.,* 2006; Pan *et al.,* 2016). The Min-Max normalization is one of the most popular methods in animal breeding and neural network studies. This normalization technique provides a linear transformation on the original data. Min-Max normalization maps the data value within a range of "0" to "1" or from "-1" to "1". Preserving the relationship between the original data is the advantage of Min-Max normalization. However, this method is highly affected by extreme values or outliers because the results can be dominated by specific large values (Jain *et al.,* 2005). In the D-Min-Max normalization, the data is scaled in [0.1-0.9] range and similarly functions to Min-Max (Han and Kamber, 2006; Shalabi *et al.,* 2006; Pan *et al.,* 2016). The D-Min-Max technique (adjusted or modified Min-Max normalization) does not perform well in case of some time series forecasting studies. It can occur a problem if any of the out-of-sample data points is out of the range (Panigrahi and Behera, 2013). In the Median method, each sample is normalized by proportioning the observations to the median of each variable. Extreme values or outliers does not affect the Median normalization. The data is scaled between "0.0- 1.0" or "-1.0- 1.0" in the Sigmoid normalization method. Sigmoid normalization method can be used for the parameters that are estimated from noisy data (Jayalakshmi and Santhakumaran, 2011). Normalized value is obtained by moving the decimal point of values of variable. The number of decimal points moved depends on the maximum absolute value of the variable (Han and Kamber, 2006). It similarly functions to the Min-Max normalization technique (Jain *et al.,* 2005; Jayalakshmi and Santhakumaran, 2011; Panigrahi and Behera, 2013).

Table 2. Normalization techniques

Normalization Techniques	Equations	Descriptions
Z-Score	$x_i{}' = \frac{x_i - \mu_i}{\sigma_i}$	μ_i: Mean value of x_i, σ_i:Standard deviation of x_i.
Min-Max	$x_i{}' = \frac{x_i - x_{imin}}{x_{max} - x_{min}}$	x_{min} and x_{max} minimum and maximum value of x_i, respectively.
D-Min-Max	$x_i' = 0.1 + (0.9 - 0.1)\frac{x_i - x_{min}}{x_{max} - x_{min}}$	x_{min} and x_{max} minimum and maximum value of x_i, respectively.
Median	$x_i{}' = \frac{x_i}{Median\ (x_i)}$	$Median\ (x_i)$: Median value of x_i.
Sigmoid	$x_i{}' = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}}$	e: Natural logarithm based.
Decimal Scaling	$x_i{}' = \frac{x_i}{10^k}$	k: The smallest integer Max ($\lvert x_n \rvert$) < 1.
Median and MAD	$x_i{}' = \frac{x_i - median}{MAD}$	Median: The median value of x_i, Median absolute deviation: Median ($\lvert x_i - median \rvert$)
Tanh-Estimator	$x_i{}' = \frac{1}{2}\left\{tanh\left(0.01\left(\frac{x_i - \mu}{\sigma}\right)\right) + 1\right\}$	μ_i and σ_i is mean and standard deviation.

$x'x_i$: Original value of x_i.

Median and Median Absolute Deviation (MAD) is one of the robust normalization techniques because of it is insensitive to outliers (Pan *et al.,* 2016; Jain and Bhandare, 2011). MAD is a measure of statistical distribution and this measure is more resilient to extreme values or outliers compared with standard deviation (Nayak *et al.,* 2014). In the Median and MAD method, the median value and the MAD value are calculated separately, and then applications are performed with the relevant formula (Jain *et al.,* 2005; Kandanaarachchi *et al.,* 2019). Tanh-Estimators normalization technique is one of the robust and efficient methods (Jain *et al.,* 2005; Nayak *et al.,* 2014), but it is not much encountered in animal breeding studies. This technique was introduced in the study published by (Hampel *et al.,* 1986). In the present study, the mathematical representations and descriptions of the mentioned normalization techniques are given in Table 2.

Accuracy of the Models

Accuracy of examined neural networks structures were determined with Adjusted Coefficient of Determination (R^2_{Adj}), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). According to these performance criteria, R^2_{Adj} with high value, RMSE and MAPE with low value is indicated as good. Mathematical expressions were given in Table 3.

Table 3. Statistical error criteria

Statistical Error Criteria	Equations
Root Mean Square Error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
Mean Absolute Percentage Error (MAPE)	$\mathrm{MAPE} = (\frac{100}{n})\sum_{i=1}^{n}\left\|\frac{(y_i-\hat{y}_i)}{y_i}\right\|$
Adjusted Coefficient of Determination (R^2_{Adj})	$R^2_{Adj} = 1 - (1 - R^2)\frac{(n-1)}{(n-p-1)}$

Where for the i^{th} record, $\hat{y}_i$: predicted value, y_i: actual value, n: number of records.

RESULTS and DISCUSSION

Various neural network architectures were evaluated and tested in order to investigate the effect of normalization techniques on the prediction of 305-day milk yield using neural networks. The normalization techniques were applied separately on data sets before training process of neural network.

305-day milk yield's normalized value was given in Figure 2. As we can see in Q-Q plot, the observed values of 305-day milk yield are distributed in similar ranges at all normalization techniques. In the original data, there were magnitude differences between the observation values of the inputs by their nature. In such cases, it can be occurred some problems related to inputs with high mathematical value such as risk of suppressing other variables and decreasing their effectiveness in neural network analysis.

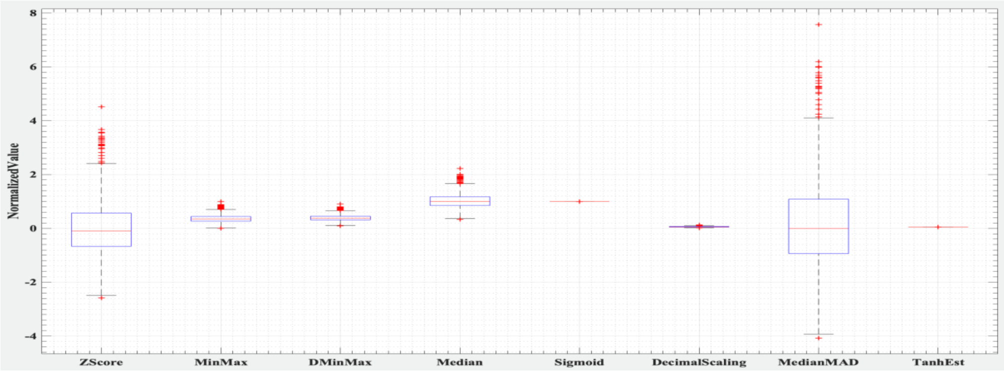

Figure 2. Q-Q plot of output's normalized value

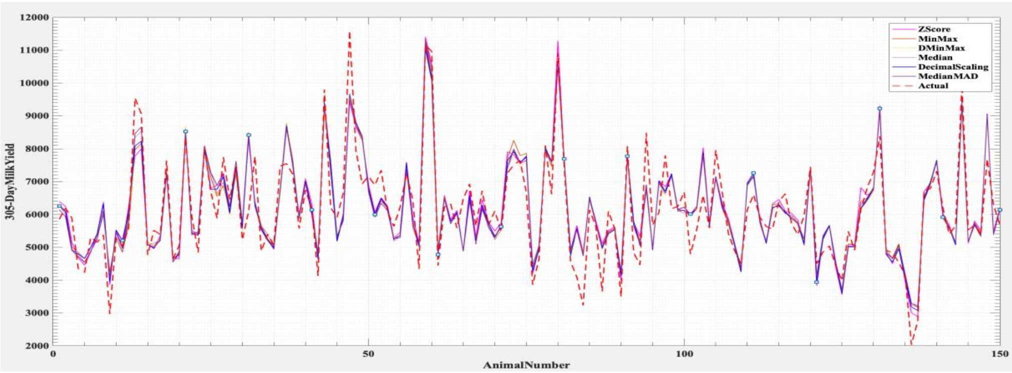

Figure 3. Observed and fitted 305-d milk yield using neural network in test phase

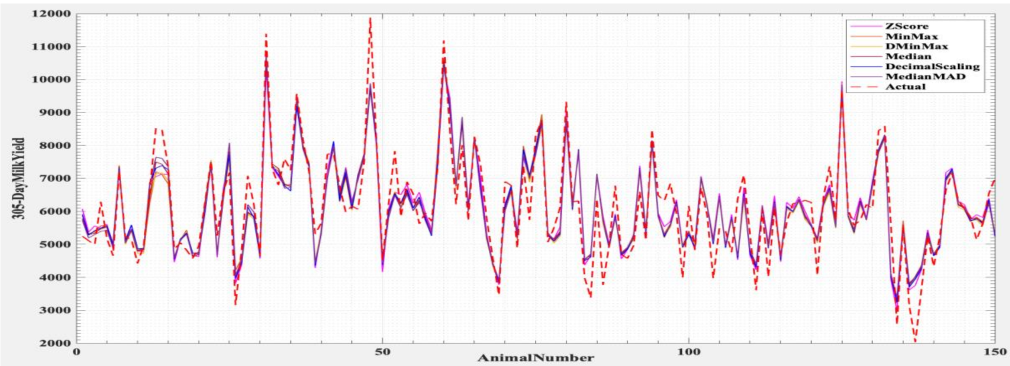

Figure 4. Observed and fitted 305-d milk yield using neural network in validation phase

Figure 3 and Figure 4 show that observed and predicted 305-day milk yield using neural network with Z-Score, Min-Max, D-Min-Max, Median, Decimal Scaling, and Median and Mean Absolute Deviation normalization techniques in test and validation phases, respectively. It can be seen that the lines representing the observed and estimated values are very close to each other in all normalization techniques.

The prediction performance of 305-day milk yield for test and validation set are shown in Table 4 and Table 5, respectively. Table 4 and Table 5 show that most successful prediction performance is obtained using BR algorithm with decimal scaling normalization technique for 305-day milk yield (R^2_{Adj}= 0. 8181, RMSE= 0.0068, MAPE= 160.42 for test set; R^2_{Adj}= 0. 8141, RMSE= 0.0067, MAPE= 114.12 for validation set). It can be observed from Table 4 and Table 5 that R^2_{Adj} values are close to each other except BFG and SCG algorithms. Also, other successful performance results were obtained as follows; SCG algorithm with decimal scaling normalization technique (R^2_{Adj} = 0. 7918, RMSE= 0.0070, MAPE= 160.82 for test set; R^2_{Adj} = 0. 8074, RMSE= 0.0068, MAPE=

114.58 for validation set), CGB algorithm with tanh-estimator normalization technique (R^2_{Adj} = 0. 7980, RMSE= 0.0072, MAPE= 144.07 for test set; R^2_{Adj} = 0. 8081, RMSE= 0.0002, MAPE= 111.82 for validation set); BFG algorithm with decimal scaling normalization technique (R^2_{Adj} = 0. 7857, RMSE= 0.0071, MAPE= 160.95 for test set; R^2_{Adj} = 0. 7534, RMSE= 0.0077, MAPE= 113.91 for validation set).

Performance criteria could not be measured with Sigmoid and Tanh-Estimator normalization techniques on BR and LM algorithms. It is thought that the formula structure of normalization techniques has an effect on this situation and also the training process for neural networks which are used in this study could not be successfully performed with these normalization techniques. Nevertheless, as a result of the calculations, it was observed that RMSE and R^2_{Adj} values were calculated parallel to each other but there were significant increases in MAPE values from time to time.

Table 4. Comparisons of performance criteria of neural network with-test set

Normalization Techniques	Back Propagation Algorithms	Transfer Functions for Test Set					
		Tanh-Sig			Log-Sig		
		R^2_{Adj}	RMSE	MAPE	R^2_{Adj}	RMSE	MAPE
Z-Score	BR	0.7975	0.4346	493.68	0.7812	0.4487	462.49
Min-Max		0.8009	0.0616	537.64	0.8070	0.0606	531.85
D-Min-Max		0.8120	0.0486	560.90	0.8051	0.0496	565.39
Median		0.8168	0.1146	1387.1	0.8151	0.1148	1396.9
Sigmoid		.	.	.	.	.	.
Decimal Scaling		0.8181	0.0068	160.42	0.8154	0.0069	160.52
Median and MAD		0.8001	0.7043	810.95	0.7950	0.7131	874.94
Tanh-Estimator		.	.	.	.	.	.
Z-Score	LM	0.7370	0.4955	545.90	0.7957	0.4347	533.37
Min-Max		0.8024	0.0603	543.01	0.7372	0.0692	535.75
D-Min-Max		0.7575	0.0540	570.52	0.7799	0.0505	555.02
Median		0.7978	0.1148	1383.0	0.7836	0.1192	1393.5
Sigmoid		.	.	.	.	.	.
Decimal Scaling		0.7895	0.0070	160.61	0.7793	0.0075	161.58
Median and MAD		0.7806	0.7377	889.44	0.7914	0.7286	927.64
Tanh-Estimator		.	.	.	.	.	.
Z-Score	SCG	0.7988	0.4311	508.69	0.7601	0.4724	600.17
Min-Max		0.7291	0.0704	551.38	0.8032	0.0597	533.90
D-Min-Max		0.8060	0.0479	566.24	0.7812	0.0504	569.79
Median		0.8018	0.1137	1404.9	0.7268	0.1351	1407.2
Sigmoid		0.1648	0.0004	1288.8	0.0421	0.0005	1288.8
Decimal Scaling		0.7918	0.0070	160.82	0.7969	0.0069	160.68
Median and MAD		0.7582	0.7726	997.44	0.6974	719.93	41830.0
Tanh-Estimator		0.0772	0.0005	143.72	-0.017	0.0006	143.87
Z-Score	CGB	0.7769	0.4564	494.01	0.7946	0.4350	506.69
Min-Max		0.7449	0.0687	539.71	0.7978	0.0611	542.40
D-Min-Max		0.7398	0.0551	545.33	0.7899	0.0495	559.39
Median		0.7388	0.1310	1388.0	0.7872	0.1182	1398.7
Sigmoid		0.7458	0.0001	1289.0	0.7418	0.0001	1289.0
Decimal Scaling		0.7134	0.0082	161.42	0.7250	0.0084	161.11
Median and MAD		0.7769	0.7464	897.29	0.6353	0.9578	1056.0
Tanh-Estimator		0.7980	0.0072	144.07	0.7211	0.0072	144.17
Z-Score	BFG	0.7616	0.4846	514.07	0.7357	0.4931	540.99
Min-Max		0.7105	0.0728	551.06	0.7352	0.0699	516.81
D-Min-Max		0.7791	0.0507	564.42	0.7846	0.0503	576.15
Median		0.7776	0.1207	1405.7	0.7746	0.1218	1408.6
Sigmoid		-0.052	0.0005	1289.5	0.2671	0.0004	1289.4
Decimal Scaling		0.7857	0.0071	160.95	0.6940	0.0085	159.75
Median and MAD		0.7804	0.7378	823.44	0.8016	0.7027	878.62
Tanh-Estimator		0.5359	0.0004	143.74	0.0653	0.0007	143.51

The increase in MAPE values is thought to be due to different normalization techniques forming different data scales. In terms of prediction accuracy with RMSE, compared to Min-Max normalization, which is very popular technique used in animal science studies, D-Min-Max normalization has a slightly better prediction.

Table 5. Comparisons of performance criteria of neural network with-validation set

Normalization Techniques	Back Propagation Algorithms	Transfer Functions for Validation Set					
		Tanh-Sig			Log-Sig		
		R^2_{Adj}	RMSE	MAPE	R^2_{Adj}	RMSE	MAPE
Z-Score	BR	0.8093	0.4224	204.13	0.8118	0.4204	195.11
Min-Max		0.8104	0.0595	266.57	0.8141	0.0588	264.16
D-Min-Max		0.8135	0.0472	284.19	0.8122	0.0472	286.71
Median		0.8129	0.1118	678.78	0.8148	0.1109	681.03
Sigmoid		.	.	.	.	.	.
Decimal Scaling		0.8141	0.0067	114.12	0.8197	0.0067	114.19
Median and MAD		0.8156	0.6828	321.39	0.8129	0.6871	322.96
Tanh-Estimator		.	.	.	.	.	.
Z-Score	LM	0.7870	0.4477	224.75	0.8108	0.4237	205.68
Min-Max		0.8085	0.0596	268.29	0.7814	0.0641	269.95
D-Min-Max		0.8005	0.0486	292.12	0.7818	0.0514	284.51
Median		0.8007	0.1171	665.24	0.8099	0.1121	687.67
Sigmoid		.	.	.	.	.	.
Decimal Scaling		0.8062	0.0068	114.40	0.8078	0.0071	115.39
Median and MAD		0.7937	0.7228	334.03	0.8100	0.6953	318.42
Tanh-Estimator		.	.	.	.	.	.
Z-Score	SCG	0.8110	0.4199	219.13	0.7863	0.4472	224.58
Min-Max		0.7771	0.0642	265.12	0.8125	0.0595	263.08
D-Min-Max		0.8072	0.0477	286.80	0.7823	0.0508	285.27
Median		0.8112	0.1120	678.20	0.7714	0.1230	669.78
Sigmoid		0.1911	0.0004	694.87	0.0437	0.0005	694.98
Decimal Scaling		0.8074	0.0068	114.58	0.8036	0.0068	114.31
Median and MAD		0.7920	0.7232	343.94	0.7475	0.8180	293.87
Tanh-Estimator		0.0656	0.0005	111.84	-0.054	0.0007	111.86
Z-Score	CGB	0.8028	0.4295	215.27	0.7965	0.4372	225.98
Min-Max		0.7935	0.0619	265.14	0.8026	0.0605	268.13
D-Min-Max		0.7420	0.0552	285.21	0.7975	0.0491	288.43
Median		0.7769	0.1219	674.15	0.7988	0.1153	686.06
Sigmoid		0.6976	0.0001	695.03	0.7042	0.0001	695.03
Decimal Scaling		0.7684	0.0074	114.59	0.7571	0.0076	113.69
Median and MAD		0.7961	0.7163	321.51	0.7155	0.8476	385.75
Tanh-Estimator		0.8081	0.0002	111.82	0.7233	0.0002	111.82
Z-Score	BFG	0.7818	0.4510	261.98	0.7829	0.4510	259.08
Min-Max		0.7530	0.0686	261.89	0.7675	0.0659	261.58
D-Min-Max		0.8070	0.0479	287.60	0.8053	0.0480	283.10
Median		0.8032	0.1146	673.25	0.7861	0.1193	676.45
Sigmoid		-0.056	0.0005	695.35	0.3186	0.0004	695.24
Decimal Scaling		0.7534	0.0077	113.91	0.7775	0.0089	113.97
Median and MAD		0.7889	0.7409	353.23	0.8056	0.6997	371.28
Tanh-Estimator		0.5652	0.0004	111.75	0.0785	0.0007	111.75

Table 4 and Table 5 show that the differences between the maximum RMSE and MAPE of normalization methods are varying from 0.0002 to 0.8476 and 111.75 to 695.35, respectively. These results mean that normalization techniques could have the possibility to change the prediction accuracy by those values.

Within the scope of applied sciences, it is seen that our study results are in harmony with other studies related to normalization techniques. Cihan *et al.* (2017) were aimed to determine the successful normalization technique for the data set by examining the effect of normalization techniques on neural network and feature selection performance in neonatal lamb diagnostics. Different from our study, it has been determined that the most successful normalization technique in diagnosing disease in neonatal lambs is sigmoid normalization. Jain *et al.* (2005) examined the effect of different score normalization techniques on the performance of a multimodal biometric system for a classification problem. In their study, min–max, Z-Score and tanh normalization techniques followed by a simple sum of scores fusion method result in a superior genuine acceptance rate than all the other normalization and fusion techniques. Also, it was determined that both Min–Max and Z-Score methods are sensitive to outliers, different from our study. Jayalakshmi and Santhakumaran (2011) aimed to propose various statistical normalization procedures to improve the classification accuracy. Different from our study, the best normalization method in the back propagation neural network model was suggested as statistical column method. Shanker *et al.* (1996) evaluated the effectiveness of two well-known transformation methods: linear transformation and statistical standardization. They indicated that the effect of data standardization on computation time and number of iterations may thus be different for other algorithms. The common denominator of our study with Shanker *et al.* (1996) is that experimental results show how data standardization methods affect neural network performance in terms of predictive accuracy, computation time and number of iterations. In the literature, there are different scientific publications on applied sciences that are compatible with the results of our study (Sola and Sevilla, 1997; Panigrahi and Behera, 2013; Nayak *et al.,* 2014; Eesa and Arabo, 2017).

CONCLUSION

Nowadays, following the technological innovation and providing effective data management is key to the efficient use of information. Milk yield estimation methods are an important source of information for the dairy cattle industry. The effective use of these methods can provide important contributions to strategic decisions and sustainable competitive advantage, in terms of animal health protection and financing. In the present study, 305-day milk yield estimation which is an important concept for dairy cattle industry, has been approached together with neural networks, which is one of the powerful prediction methods. In the study, it is aimed to increase the prediction accuracy with the improvements in neural network performance. In this way, it is aimed to contribute to the users to be more successful in future production and management planning. For his purpose, hundreds of neural network architectures based on eight different normalization methods, two different activation functions, five different back propagation algorithms, and different learning parameters based on heuristics and standard numerical optimization techniques are experimentally investigated for the network optimization. Analysis results show that with the use of different normalization techniques, different performance values have been obtained in various neural network architectures. According to the results of this study, effect of the distribution of output in a training set with normalization varies with various factors, such as the neural network architectures. It's deduced that the best training algorithm is Bayesian Regularization with decimal scaling normalization that attains more than 80% prediction accuracy. The most successful prediction value was obtained with this optimized neural network structure for 305-day milk yield. Neural networks approach can be improved with other training algorithms and learning parameters or data preprocessing elements. At the same time, other production and reproduction traits of dairy cattle can be used in terms of the improvement of prediction 305-day milk yield model. Results of analysis show that applying different normalization techniques affect the performance and the distribution of outputs influences the learning process of neural network. The magnitude of the effects varied with the type of back propagation algorithms, activation functions and network's architectural structure. In the focus of this study, it is aimed to present that different results can be calculated with different normalization techniques in the neural network analysis process in order to improve the prediction accuracy. The results of this study are intended to be useful for animal breeders and provide information about economic traits of importance in dairy enterprise.

DECLARATION OF COMPETING INTEREST

The authors declare that there is no conflict of interest.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Asli Akilli: Methodology, conceptualization, literature review, formal analysis in Matlab, writing and editing of manuscript, visualization.
Hulya Atil: Review and editing of manuscript, interpretation of the analysis results.

REFERENCES

Akıllı A (2019). *Analysis of agricultural data with multivariate nonlinear fuzzy regression method*. PhD Thesis, Kırşehir Ahi Evran University, Institute of Science, Department of Zootechnics, p. 249, Kırşehir, Turkey (in Turkish).

Atıl H and Akıllı A (2016). Comparison of artificial neural network and K-means for clustering dairy cattle. *International Journal of Sustainable Agricultural Management and Informatics, 2 (1): 40-52.*

Brown-Brandl TM, Jones DD and Woldt WE (2005). Evaluating modelling techniques for cattle heat stress prediction. *Biosystems Engineering, 91 (4): 513-524.*

Chen LJ, Cui LY, Xing L and Han LJ (2008). Prediction of the nutrient content in dairy manure using artificial neural network modeling. *Journal of Dairy Science, 91: 4822-4829.*

Cihan P, Kalıpsız O and Gökçe E (2017). Effect of normalization techniques on artificial neural network and feature selection performance in animal disease diagnosis. *Electronic Turkish Studies, 12 (11): 59-70.*

Dong R and Zhao G (2014). The use of artificial neural in vitro rumen methane production using the CNCPS carbohydrate fractions as dietary variables. *Livestock Production Science, 162: 159-167.*

Dongre VB, Gandhi RS, Singh A and Ruhil AP (2012). Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. *Livestock Science, 147: 192–197.*

Edriss MA, Hosseinnia P, Edrisi M, Rahmani HR and Nilforooshan MA (2008). Prediction of second parity milk performance of dairy cows from first parity information using artificial neural network and multiple linear regression methods. *Asian Journal of Animal and Veterinary Advances, 3 (4): 222-229.*

Eesa AS and Arabo WK (2017). Normalization method for backpropagation: A comparative study. *Science Journal of University of Zakho, 5 (4): 319-323.*

Gandhi RS, Monalisa D, Dongre VB, Ruhil AP, Singh A and Sachdeva GK (2012). Prediction of first lactation 305-day milk yield based on monthly test day records using artificial neural networks in Sahiwal cattle. *Indian Journal of Dairy Science, 65 (3): 229-233.*

Gandhi RS, Raja TV, Ruhil AP and Kumar A (2010). Artificial neural network versus multiple regression analysis for prediction of lifetime milk production in Sahiwal cattle. *Journal of Applied Animal Research, 38 (2): 233-237.*

Gorgulu O (2012). Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *South African Journal of Animal Science, 42 (3): 280-287.*

Grzesiak W, Lacroix R, Wójcik J and Blaszczyk P (2003). A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science, 83 (2): 307-310.*

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). *Computers and Electronics in Agriculture, 54: 69-83.*

Hampel FR, Ronchetti EM, Rousseeuw PJ and Stahel WA (1986). Robust statistics: The approach based on influence functions. *John Wiley and Sons:* Canada.

Han J and Kamber M (2006). Data mining: Concepts and techniques. *2nd ed, Elsevier*: San Francisco, USA.

Hassan KJ, Samarasinghe S and Lopez- Benavidest MG (2009). Use of neural networks to detect minor and major pathogens that cause bovine mastitis. *Journal of Dairy Science, 92: 1493-1499.*

Hosseinia P, Edrisi M, Edriss MA and Nilforooshan MA (2007). Prediction of second parity milk yield and fat percentage of dairy cows based on first parity information using neural network system. *Journal of Animal and Poultry Sciences, 7 (21): 3274-3279.*

Ince D and Sofu A (2013). Estimation of lactation milk yield of Awassi sheep with artificial neural network modeling. *Small Ruminant Research, 113 (1): 15-19.*

Jain A, Nandakumar K and Ross A (2005). Score normalization in multimodal biometric systems. *Pattern Recognition, 38 (12): 2270-2285.*

Jain YK and Bhandare SK (2011). Min max normalization based data perturbation method for privacy protection. *International Journal of Computer & Communication Technology, 2 (8): 45-50.*

Jayalakshmi T and Santhakumaran, A (2011). Statistical normalization and back propagation for classification. *International Journal of Computer & Communication Technology, 3 (1): 1793-8201.*

Kandanaarachchi S, Muñoz MA, Hyndman RJ and Smith-Miles K (2019). On normalization and algorithm selection for unsupervised outlier detection. *Data Mining and Knowledge Discovery, 34 (2): 309-354.*

Karadas K, Tariq M, Tariq MM and Eyduran, E (2017). Measuring predictive performance of data mining and artificial neural network algorithms for predicting lactation milk yield in indigenous Akkaraman sheep. *Pakistan Journal of Zoology, 49 (1): 1-7.*

Kominakis AP, 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. *Computers and Electronics in Agriculture, 35 (1): 35-48.*

Kong LN, Li JB, Li RL, Zhao XX, Ma YB, Sun SH and Zhong JF (2018). Estimation of 305-day milk yield from test-day records of Chinese Holstein cattle. *Journal of Applied Animal Research, 46 (1): 791-797.*

Mostert BE, Theron HE, Kanfer FHJ and van Marle Koster E (2006). Comparison of breeding values and genetic trends for production traits estimated by a lactation model and a fixed regression test- day model. *South African Journal of Animal Science, 36: 71–78.*

Murphy MD, O'Mahony MJ, Shalloo L, French P and Upton J (2014). Comparison of modelling techniques for milk-production forecasting. *Journal of Dairy Science, 97 (6): 3352-3363.*

Nayak SC, Misra BB and Behera HS (2014). Impact of data normalization on stock index forecasting. *International Journal of Computer Information Systems and Industrial Management Applications, 6: 357-369.*

Negnevitsky M (2002). Artificial intelligence, a guide to intelligent systems. 2nd ed, *Pearson Education*: Harlow.

Njubi DM, Wakhungu JW and Badamana MS (2010). Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. *Tropical Animal Health and Production, 42 (4): 639-644.*

Njubi DM, Wakhungu J and Badamana MS (2009). Milk yield prediction in Kenyan Holstein-Friesian cattle using computer neural networks system. *Livestock Research for Rural Development, 21 (4).*

Oztemel E (2002). Yapay sinir ağları. *Papatya Yayıncılık*, İstanbul, Turkey.

Pan J, Zhuang Y and Fong S (2016). *The impact of data normalization on stock market prediction: Using SVM and technical indicators.* In International Conference on Soft Computing in Data Science, 2nd Ed: Berry MW, Mohamed AH, Yap BW, *Springer*: Singapore, 72-88.

Panigrahi S and Behera HS (2013). Effect of normalization techniques on univariate time series forecasting using evolutionary higher order neural network. *International Journal of Engineering and Advanced Technology, 3 (2): 280-285.*

Ruhil AP, Gandhi RS, Monalisa D, Behra K and Raja TV (2011). *Prediction of lactation yield based on partial lactation records using artificial neural networks.* In Proc. 5th National Conference on Computing for Nation Development, INDIACom-2011.

Russell SJ and Norvig P (2016). Artificial intelligence: A modern approach. *Pearson Education Limited:* Malaysia.

Salehi F, Lacroix R and Wade KM (1998). Effects of learning parameters and data presentation on the performance of backpropagation networks for milk yield prediction. *Transactions of the ASAE, 41 (1): 253.*

Sanzogni L, and Kerr D (2001). Milk production estimates using feed forward artificial neural networks. *Computers and Electronics in Agriculture, 32 (1): 21-30.*

Savegnago RP, Nunes BN, Caetano SL, Ferraudo AS, Schmidt GS, Ledur MC and Munari DP (2011). Comparison of logistic and neural network models to fit to the egg production curve of White Leghorn hens. *Poultry Science, 90 (3): 705-711.*

Shahinfar S, Mehrabani-Yeganeh H, Lucas C, Kalhor A, Kazemian M and Weigel KA (2012). Prediction of breeding values for dairy cattle using artificial neural networks and neuro-fuzzy systems. *Computational and Mathematical Methods in Medicine, 2012: 127130.*

Shalabi AL, Shaaban Z and Kasasbeh B (2006). Data mining a preprocessing engine. *Journal of Computer Science, 2 (9): 735-739.*

Shanker M, Hu MY and Hung MS (1996). Effect of data standardization on neural network training. *Omega, 24 (4): 385-397.*

Sharma AK, Sharma RK and Kasana HS (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 Computing & Applications, 15 (3-4): 359-365.*

Sharma AK, Sharma RK and Kasana HS (2007). Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing,* 7: 1112-1120.

Sola J and Sevilla J (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science, 44 (3): 1464-1468.*

Tahmoorespur M, Hosseinnia P, Teimurian M and Aslaminejad AA (2012). Predictions of 305-day milk yield in Iranian Dairy cattle using test-day records by artificial neural network. *Indian Journal of Animal Sciences, 82 (5): 511-516.*

Takma Ç, Atıl H and Aksakal V (2012). Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. *Journal of the Faculty of Veterinary Medicine, Kafkas University, 18 (6): 941-944.*

Torres M, Hervás C and Amador F (2005). Approximating the sheep milk production curve through the use of artificial neural networks and genetic algorithms. *Computers & Operations Research, 32 (10): 2653-2670.*

Yang XZ, Lacroix R and Wade KM (2000). Investigation into the production and conformation traits associated with clinical mastitis using artificial neural networks. *Canadian Journal of Animal Science, 80: 415-426.*

Zhang G, Patuwo BE and Hu MY (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting,14 (1): 35-62.*