Predicting the body weight of Hereford cows using machine learning

A N Ruchay\textsuperscript{1,2}, V I Kolpakov\textsuperscript{1}, V V Kalschikov\textsuperscript{2}, K M Dzhulamanov\textsuperscript{1} and K A Dorofeev\textsuperscript{2}

\textsuperscript{1}Federal Research Centre of Biological Systems and Agro-technologies of the Russian Academy of Sciences, Orenburg, Russia
\textsuperscript{2}Department of Mathematics, Chelyabinsk State University, Chelyabinsk, Russia

E-mail: ran@csu.ru, vkolpakov056@yandex.ru, vkalschikov@gmail.com, kinispai.d@yandex.ru, kostuan1989@mail.ru

Abstract. Various machine learning algorithms have been used to model and predict the body weight of Hereford cows. The traditional linear regression model and various machine learning algorithms have been used to develop models for the prediction of the body weight of Hereford cows. The dependent variables include body weight and independent variables include withers height, hip height, chest dept, chest width, width in maclocks, sciatic hill width, oblique length of the body, oblique rear length, chest girth, metacarpus girth, backside half-girth, and age measurements of 1500 cows aged 2–6 years of age. The performance of the models is assessed based on evaluation criteria of the coefficient of determination, the root mean squared error, the mean absolute error, the mean absolute percentage error. We used a concept of splitting data into training, testing and validation datasets to provide a robust method for modelling and predicting. The RandomForestRegressor algorithm was found to provide the best results for training and testing datasets. It was concluded that machine learning algorithms may provide better results than the traditional models and may help researchers choose the best predictors for body weight of animals.

1. Introduction

Body weight is an important measure of stock performance, which provides an informative measure for feeding, health care, breeding, and selection of stocks. Also, measurement of live weight of stock is one of the most important production tools available to farmers, playing a role in nutrition, fertility management, health and marketing [1]. Body weight prediction can be based on automatically measured morphological traits using the 2D vision system [2, 3] and the 3D vision system [4–7].

Most past studies have employed multiple linear regression analysis for predicting the body weight of animals. However, these traditional methods are inadequate for predicting [8]. Recently, a few researchers have successfully applied various machine learning algorithms for the prediction of live body weight of animals using morphological measures [3, 8–11]. These methods aim to map body weight from morphological measures of animals. These studies have reported the potential of machine learning algorithms in accurately predicting the nonlinear relation between body weight and morphological and biometrical traits of animals [8].
In this study various machine learning methods for developing a body weight prediction model for Hereford cows have been investigated. No studies in the literature, to our knowledge, have reported on the prediction of body weight of Hereford cows by exploiting the combination of machine learning methods.

This study aimed to determine the best machine learning methods to predict the body weight of Hereford cows using various morphological. Also as part of the machine learning approach we have used a concept of splitting data into training, testing and validation datasets to provide a robust method for modelling and predicting. The training of the model is carried out using a training dataset. Usually, at least 70% of the data is used for training. The parameters of the model are then tuned using validation set as part of training dataset. Usually, at least 20% of the training set is used for validation. Finally, the actual predictive power of the model is verified using a testing set.

In this study the traditional linear regression model (LinearRegression [12]) and various machine learning algorithms [13, 14] have been employed for modelling and predicting body weight of Hereford cows from body size measurements (withers height, hip height, chest dept, chest width, width in maclocks, sciatic hill width, oblique length of the body, oblique rear length, chest girth, metacarpus girth, backside half-girth) and age (full years).

2. Materials and methods

2.1. Data collection

This study utilizes data from 1500 Hereford cows kept in private farms with top dressing with concentrated feed, Nizhny Novgorod region, Russia, and in private farms with grain fattening at the open type site, Voronezh region, Russia.

The dependent variable body weight and independent variables such as withers height, hip height, chest dept, chest width, width in maclocks, sciatic hill width, oblique length of the body, oblique rear length, chest girth, metacarpus girth, backside half-girth, and age were measured for cow aged 2 to 6 years using tailor tape and weigh balance.

2.2. Ethics in animal care

This study was conducted according to the guidelines issued by the Animal Care and Use Committee of Federal Research Center of Biological Systems and Agro-Technologies of the Russian Academy of Sciences. The dataset collection and all other procedures in this study were carried out without restraining the cows. This study does not require approval from the Animal Care and Use Committee that excludes ecology monitoring or livestock sciences from requiring approval.

2.3. Machine learning algorithms

We use the following regression algorithms from the SciKit-Learn (SKlearn) library [27]: Bayesian ARD (ARDRegression), AdaBoost (AdaBoostRegressor), Bagging (BaggingRegressor), Bayesian ridge (BayesianRidge), Decision tree (DecisionTreeRegressor), Elastic Net (ElasticNetCV), Extra trees (ExtraTreesRegressor), Gaussian process (GaussianProcessRegressor), Gradient Boosting (GradientBoostingRegressor), Histogram-based Gradient Boosting Regression Tree (HistGradientBoostingRegressor), Linear regression Huber model (HuberRegressor), Isotonic regression model (IsotonicRegression), Regression based on k-nearest neighbors (KNeighborsRegressor), Kernel ridge (KernelRidge), Lasso linear model with cross-validation (LassoCV), Ordinary least squares Linear Regression (LinearRegression), Linear Support Vector Regression (LinearSVR), Multi-layer Perceptron (MLPRegressor), Cross-validated Orthogonal Matching Pursuit model (OrthogonalMatchingPursuitCV), Partial least squares regression (PLSRegression), Passive Aggressive (PassiveAggressiveRegressor), RANSAC (RANdom SAmple Consensus) (RANSACRegressor), Regression based on neighbors within a fixed radius (RadiusNeighborsRegressor), Random Forest (RandomForestRegressor), Ridge regression with cross-
validation (RidgeCV), Linear model fitted based on Stochastic Gradient Descent (SGDRegressor), Epsilon-Support Vector (SVR), Theil-Sen Estimator (TheilSenRegressor).

However, some of the algorithms during the test gave poor results and their results are not presented in this article. For evaluating models we present results of the following regression algorithms: ExtraTreesRegressor, RandomForestRegressor, KNeighborsRegressor, LinearRegression, SVR, GradientBoostingRegressor, DecisionTreeRegressor, AdaBoostRegressor, MLPRegressor, RidgeCV, LassoCV, LassoLarsCV, OrthogonalMatchingPursuitCV, BayesianRidge, TheilSenRegressor, HuberRegressor.

The dataset was initially partitioned randomly into two parts, the training (70%) and testing (30%) datasets. Also 2% of the training dataset is used for validation.

2.4. Model evaluation
Different evaluation criteria have been employed to assess the performance of the models developed in this study for modelling and predicting the body weight of cow.

We consider a variety of commonly used evaluation measures in this study. We use the coefficient of determination ($R^2$), the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) as measures of evaluation quality. They are defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - f_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - f_i)^2},$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - f_i|,$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - f_i}{y_i} \right|,$$

where $n$ is a number of samples of dataset, $\bar{y}$ is the mean of all known value of body weight, $y_i$, $i = 1, \ldots, n$ is a known value of body weight, and $f_i$, $i = 1, \ldots, n$ is a predicted value of body weight.

3. Results
Table 1 shows the results of various evaluation measures used to evaluate a model’s performance on both training and testing datasets. It can be noticed from the results of the table that only ExtraTreesRegressor, RandomForestRegressor, KNeighborsRegressor, SVR, GradientBoostingRegressor algorithms can be used to model the body weight of cow, however the RandomForestRegressor algorithm gave the best result on all evaluation measures for both training and testing datasets.
4. Conclusion
This study employed the traditional linear regression model and various machine learning algorithms to predict the body weight of Hereford cows using various body measures (withers height, hip height, chest dept, chest width, width in maclocks, sciatic hill width, oblique length of the body, oblique rear length, chest girth, metacarpus girth, backside half-girth) and age. Using various evaluation measures, we found strong evidence of better performance for machine learning algorithms. The RandomForestRegressor algorithm was found to provide more accurate predictions of body weight, outperforming the traditional linear model. Based on the results of the present study, we conclude that the RandomForestRegressor algorithm can be used to model and predict body weight of Hereford cows. The findings of this study may help researchers and practitioners to adopt the latest machine learning algorithms for accurate prediction of body weight using various morphological traits and other factors.

References
[1] Dzhulamanov K, Gerasimov N, Ruchay A, Kolpakov V and Dzhulamanov E 2019 The assessment of morphological features in hereford cattle IOP Conf. Ser. Earth and Environmental Sci. 341 012062
[2] Tasdemir S, Urkmez A and Inal S 2011 Determination of body measurements on the holstein cows using digital image analysis and estimation of live weight with regression analysis Comp. and Electr. in Agricult. 76 189–97
[3] Weber V, Weber F, Gomes R et al 2020 Prediction of girolando cattle weight by means of body measurements extracted from images Revista Brasileira de Zootecnia 49
[4] Kuzuhara Y, Kawamura K, Yoshitoshi R et al 2015 A preliminarily study for predicting body weight and milk properties in lactating holstein cows using a three-dimensional camera system Comp. and Electr. in Agricult. 111 186–93
[5] Song X, Bokkers E, van der Tol P, Koerkamp P G and van Mourik S 2018 Automated body weight prediction of dairy cows using 3-dimensional vision J. of Dairy Sci. 101 4448–59
[6] Ruchay A, Dorofeev K, Kalschikov V, Kolpakov V and Dzhulamanov K 2019 A depth camera-based system for automatic measurement of live cattle body parameters *IOP Conf. Ser. Earth and Environmental Sci.* **341** 012148

[7] Ruchay A, Dorofeev K, Kalschikov V, Kolpakov V and Dzhulamanov K 2019 Accurate 3D shape recovery of live cattle with three depth cameras *IOP Conf. Ser. Earth and Environmental Sci.* **341** 012147

[8] Huma Z and Iqbal F 2019 Predicting the body weight of Balochi sheep using a machine learning approach *Turkish J. of veter. and animal sci.* **43** 500–6

[9] Hempstalk K, Mcparland S and Berry D 2015 Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows *J. of dairy sci.* **98**

[10] Miller G A, Hyslop J J, Barclay D, Edwards A, Thomson W and Duthie C A 2019 Using 3D imaging and machine learning to predict liveweight and carcass characteristics of live finishing beef cattle *Frontiers in Sustainable Food Syst.* **3** 30

[11] Milosevic B, Cric S, Lalic N et al 2019 Machine learning application in growth and health prediction of broiler chickens *World’s Poultry Sci. J.* **75** 401–10

[12] Pedregosa F, Varoquaux G, Gramfort A et al 2011 Scikit-learn: Machine learning in Python *J. of Machine Learn. Res.* **12** 2825–30

[13] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Springer)

[14] MacKay D J C 2003 *Information Theory, Inference & Learning Algorithms* (USA: Cambridge Univer. Press) ISBN 0521642981