International Journal of Livestock Research, Vol. 12 (2) Feb'2022
http://www.ijlr.org
eISSN: 2277-1964

Original Research

Prediction Of Egg Weight from Egg Quality Characteristics by Using Regression Analysis Methods in White Leghorn Chicken

Arjun Alapatt[1], J. K. Chaudhary[2]*, N. Shyamsana Singh[3], T. C. Tolenkhomba[4], Girin Kalita[5], Jagan Mohanarao G.[6]

[1]Department of Animal Genetics and Breeding, College of Veterinary Sciences and Animal Husbandry, Selesih, Aizawl, Mizoram, INDIA
[2]Assistant Professor, Dept. of Animal Genetics and Breeding, College of Veterinary Sciences and Animal Husbandry, Selesih, Aizawl, Mizoram, INDIA
[3]Associate Professor and Head, Dept. of Animal Genetics and Breeding, College of Veterinary Sciences and Animal Husbandry, Selesih, Aizawl, Mizoram, INDIA
[4]Assistant Professor (SG), Dept. of Animal Genetics and Breeding, College of Veterinary Sciences and Animal Husbandry, Selesih, Aizawl, Mizoram, INDIA
[5]Associate Professor, Dept. of Livestock Production and Management, College of Veterinary Sciences and Animal Husbandry, Selesih, Aizawl, Mizoram, INDIA
[6]Assistant Professor, Dept. of Veterinary Physiology and Biochemistry, College of Veterinary Sciences and Animal Husbandry, Selesih, Aizawl, Mizoram, INDIA

*Corresponding Author: vetjitu@gmail.com

How to cite this paper:
Alapatt, A., Chaudhary, J., Singh, N., Tolenkhomba, T., Kalita, G., & Jagan, M. (2022). Prediction Of Egg Weight from Egg Quality Characteristics by Using Regression Analysis Methods in White Leghorn Chicken. International Journal of Livestock Research, 12(2), 40-48. doi: 10.5455/ijlr.20210929030220

Received : Sep 29, 2021
Accepted : Feb 16, 2022
Published : Feb 28, 2022

Abstract

The study was aimed at evaluating the performance of regression methods such as Multiple Linear Regression (MLR), Ridge Regression (RR), LASSO regression, and Elastic net in the prediction of egg weight from various egg quality characteristics such as shape index (SI), yolk height (YH), yolk index (YI), albumen height (AH), Haugh unit (HU), albumen index (AI), yolk ratio (YR), albumen ratio (AR), shell weight (SW) and shell thickness (ST). For this, 100 white leghorn eggs were collected and egg quality parameters were recorded. In order to compare the predictive performances of the assigned methods, the goodness of fit criteria such as the coefficient of determination (R2), Adjusted R2, Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE) were utilized. The highest Adj. The R2 value was obtained for the Elastic net and MLR methods. Considering the multicollinearity existent, the Elastic net was identified as the best performing model.

Keywords: Egg weight, Egg quality traits, Elastic net, LASSO, Multiple Linear Regression, Ridge Regression, White Leghorn

Introduction

The avian egg primarily serves the important biological function of reproduction by conserving the multiplex of nutrients required for the growth and development of the embryo. Assessment of egg quality is done using different egg quality parameters and among those, egg weight is perhaps the most important determinant of egg quality for both table eggs and hatching eggs, as it has remarkable influence over the nutrient content of eggs and the weight of day-old chicks (Khan *et al*., 2004). Thus, variation is observed in the proportion of different components of the egg pursuant to changes in the overall weight of the egg. Also, significant variation is seen in egg weight between the various strains and varieties of poultry. As egg weight is the foremost criterion grading of eggs, adequate understanding and predicting of egg weight can provide economic benefits for producers and help in improving the methods for selecting eggs for use in reproduction (Faridi *et al.*, 2013; Okoro *et al.*, 2017). Several statistical approaches are available that can be effectively used to reveal such causal nexus in the field of animal science. These include linear regression analysis (simple and multiple), multiple regression analysis using factor analysis scores, multiple regression analysis using principal component analysis scores, Path Analysis, Regression Tree Analysis (Khan *et al.*, 2014). A promising alternative to these traditional approaches is the various data mining algorithms. Data mining makes use of computer-based methods to unveil valuable information from complex data (Kantardzic, 2011).

Considering the earlier studies and their advantages, such methods can be an effective tool in the classification of eggs, existing in egg quality criteria instead of traditional regression methods. Hence, the aim of this present investigation will be to predict Egg weight (EW) from different egg quality parameters such as, Shell Weight (SW), Shell Thickness (ST), Shape Index (SI), Yolk Height (YH), Albumin Height (AH), Yolk Index (YI), Albumin Index (AI), Haugh Unit (HU), Yolk Ratio (YR) and Albumin Ratio (AR) in White Leghorn chicken by means of Multiple Linear Regression (MLR), Ridge Regression (RR), LASSO, Elastic net, Regression Tree Method (based on CART algorithm) and Random Forest regression and to compare the efficacy of the methods. A model for predicting egg weight based on egg quality standards has more potential for improving egg quality standards.

Review of Literature

The usability of the CART algorithm for determining egg quality characteristics influencing fertility in Japanese quail eggs by Celik *et al.*, 2016. The OLS, Ridge, Lasso, and Elastic net methods were evaluated in predicting quality characteristics in Japanese quail by Ciftsuren and Akkol, 2018. It was concluded that the LASSO was the best model with variable selection. Prediction of egg weight from external egg quality traits by multiple linear regression and regression tree methods by Salgado *et al*., 2021.

Materials and Methods

The materials used in the present study were 100 eggs of White Leghorn chicken obtained from Poultry Farm, College of Veterinary Sciences and Animal Husbandry, Selesih. Egg weight (EW), Shell Weight (SW), Shell Thickness (ST), Yolk Height (YH), and Albumin Height (AH) were recorded on the obtained eggs with eggs collected daily. Shape Index (SI), Yolk Index (YI), Albumen Index (AI), Haugh unit (HU), Yolk Ratio (YR), and Albumen ratio (AR) were calculated using appropriate formulas. EW was used s the response variable and the remaining ten variables as predictors.

In Multiple Linear Regression (MLR) the estimation of parameters is performed using the ordinary least squares method (OLS). Here the values of the coefficients *β0, β1, ……, βp* are chosen in order to minimise the RSS value (Sum of Squared Residuals).

$$\text{RSS} = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 = \sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_{i1} - \widehat{\beta_2}x_{i2} - \cdots - \widehat{\beta_p}x_{ip})^2$$

Ridge regression, however provides a solution for addressing the issue of multicollinearity without removing any independent variables from the initial set (McDonald, 2009). Instead of forcing some of the coefficients to vanish, in RR method, provided *k* identical predictors, the method would yield identical coefficients that are *(1/k)*th the size which any one of the explanatory variables would get if it was fit individually.

In LASSO method, the lasso penalty presumes that only a small subgroup of coefficients is larger (nonzero) and that many coefficients are close to zero. This method obtains the β coefficients by solving the optimization problem given below:

$$\hat{\beta}_{LASSO} = \arg\min_{\beta}\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{i=j}^{k} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{k}|\beta_j|\}$$

Where, $l_1 = \sum_{j=1}^{p}|\beta_j|$ is called LASSO penalty function and here the l_1 penalty is the least squares fit and some of the components of $\hat{\beta}_{LASSO}$ are shrunk to zero. Lasso identifies and excludes those variables that are found irrelevant in the prediction thereby minimizing the model complexity and the prediction error by selecting an optimal λ (Ranstam and Cook, 2018). To evade the instability observed in lasso solution paths when there are highly correlated predictors, the Elastic net (EN) was suggested to analyse high dimensional datasets (Zou and Hastie, 2005). This method makes use of a mixture of ridge (l_2) and LASSO (l_1) penalties and it can be formulated as follows:

$$\hat{\beta}_{EN} = \left(1 + \frac{\lambda_2}{n}\right)\{\arg\min_{\beta}(\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2 + \lambda_2\sum_{j=1}^{k}\beta_j{}^2 + \lambda_1\sum_{j=1}^{k}|\beta_j|)\}$$

CART divides a set of data into segments that are as homogeneous as possible in terms of a single response variable. (Lahmann and Kottner, 2011). A regression tree can be considered as a kind of additive model of the form:

$$m(x) = \sum_{i=1}^{n} k_i \times I(x \in D_i)$$

Where k_i are constants; *I (.)* is an indicator function returning 1 if its argument is true and 0 otherwise; D_i are disjoint partitions of the training data D such that $\bigcup_{i=1}^{n} D_i = D$ and $\bigcap_{i=1}^{n} D_i = \phi$. Random forest regression and classification models works by fitting an ensemble of decision tree models to a dataset. Random forests algorithm fit separate decision trees to a predefined number of bootstrapped data sets. The predicted value for a continuous response is the mean/average fitted response from all the individual trees resulting from each bootstrapped sample.

Result and Discussion

Table 1: Descriptive statistics of egg quality characteristics in White leghorn chicken

Predictor	Mean	Std. Dev	Std. Error	Min	Max
Egg weight (EW) (g)	57.26	5.38	0.54	47.40	71.84
Shape Index (SI) (%)	74.04	4.00	0.40	64.07	81.64
Yolk Height (YH) (mm)	14.73	1.83	0.18	8.40	18.54
Yolk Index (YI) (mm)	35.49	5.27	0.53	16.55	46.37
Albumen Height (AH) (mm)	4.80	1.24	0.12	2.75	10.65
Haugh Unit (HU)	66.09	10.02	1.00	42.86	102.76
Albumen Index (AI) (%)	5.43	2.06	0.21	2.79	17.04
Yolk Ratio (YR)	31.00	2.59	0.26	25.91	38.46
Albumen Ratio (AR)	59.38	2.74	0.27	51.33	65.46
Shell Weight (SW) (g)	5.54	0.95	0.09	3.63	9.00
Shell Thickness (ST) (mm)	0.35	0.04	0.004	0.27	0.47

The analysis performed on data on egg quality characteristics in White leghorn eggs by applying multiple linear regression method showed a significant F-test (p-value: < 0.001). As interpretable from Table 2, in general, out of the ten explanatory variables identified in the present study, six are found to have a significant influence in determining the egg weight in white leghorn- namely, AH, HU, AI, YR, AR, SW (p<0.001).

Table 2: Estimated parameters and significance levels obtained using multiple linear regression analysis.

Predictor	Coefficient	SE of Coefficient	t-value	P-value
Constant	57.398	0.239	239.789	<0.001***
SI	-0.285	0.284	-1.004	>0.05
YH	0.521	0.882	0.591	>0.05
YI	0.102	0.840	0.121	>0.05
AH	16.815	2.258	7.448	<0.001 ***
HU	-7.902	1.262	-6.262	<0.001 ***
AI	-8.375	1.425	-5.879	<0.001 ***
YR	3.757	0.997	3.770	<0.001 ***
AR	4.172	0.965	4.324	<0.001 ***
SW	2.601	0.716	3.631	<0.001 ***
ST	0.074	0.293	0.254	>0.05

*Significance codes: 0'***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1;*

The estimated variance inflation factor for the different predictors in the model is shown in Table 3:

Table 3: Estimated variance inflation factor for all predictor variables in MLR

Predictor	**SI**	**YH**	**YI**	**AH**	**HU**	**AI**	**YR**	**AR**	**SW**	**ST**
VIF	1.38	13.39	12.12	87.68	27.40	34.91	17.09	16.01	8.82	1.47

Since VIF is a measure of multicollinearity in the model, the high values of VIF (more than 10) obtained for some of the independent variables is a sign of the possible high correlations existent between the predictors used. The optimum lambda value automatically generated by the fitted model was **0.1258925.** The alpha value used for lasso regression is 1. The best lambda, came out to be **lambda= 0.7943282**. The properties of ridge and lasso regression are combined in elastic net regression. The optimal values determined for the model were **alpha= 0.6422 and lambda=0.0511**. The estimated coefficients obtained using the MLR, ridge, LASSO and EN methods in the regression analyses for egg weight are given in table 4.

Table 4: Estimated coefficients using MLR, Ridge, Lasso and Elastic net methods

Predictor	**MLR**	**RIDGE**	**LASSO**	**ELASTIC NET**
Intercept	57.398	57.398	57.398	57.398
SI	-0.285	-0.484	-0.034	-0.334
YH	0.521	2.086	1.476	1.308
YI	0.102	-0.728	.	-0.374
AH	16.815	2.901	.	8.897
HU	-7.902	-1.715	.	-4.490
AI	-8.375	-1.439	.	-4.257
YR	3.757	2.124	.	4.212
AR	4.172	2.872	0.219	4.762
SW	2.601	3.319	2.056	3.736
ST	0.074	-0.256	.	-0.082

A regression tree based on CART was constructed using egg weight as response variable and ten different egg quality characteristics identified as explanatory variables. The rpart package in R statistical software was applied and the rpart.plot() function was used for visualization of the tree. The regression tree obtained is depicted in Figure 1. The SW, YH, SI and AH were used in the tree building. It was revealed that the SW is the most crucial predictor influencing the egg weight in white leghorn eggs.

Node 1 otherwise called the root node predicted an overall egg weight of 57.249 g (S=5.357). On the basis of SW, Node 1 was partitioned into Node 2 (eggs with SW<6.2 g) and node 3 (eggs with SW≥6.2 g). Predicted egg weight for node 2 was 55.762 g (S=4.537) and that for node 3 was 63.994 g (S=3.184). Node 2 was further split into nodes 4 and node 5 on the basis of YH. Eggs having YH<14 mm(n=22) and those that possess YH≥14 mm(n=37).

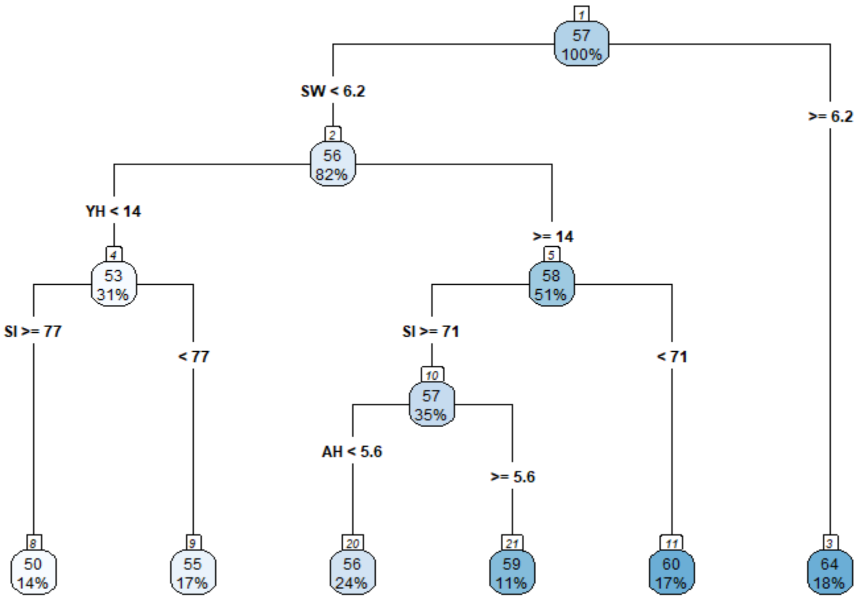

Figure 1: Regression Tree Diagram-CART Algorithm for White leghorn eggs

Nodes 4 and 5 had predicted egg weights of 52.516 g (S=3.499) and 57.693 g(S=3.969) respectively. In accordance with SI, a further split of node 4 was carried out. Node 8 included eggs possessing SI≥77.26 (n=10) and those possessing SI<77.26(n=12) were grouped in node 9. The predicted weights for eggs in nodes 8 and 9 were 50.028 g (S=1.549) and 54.589 g (S=3.332) respectively. Node 5 was further partitioned based on the SI into nodes 10 and 11. Node 10 included eggs with SI≥71.32(n=25) and those possessing SI<71.32 were grouped under node 11(n=12). Nodes 10 and 11 predicted mean EW of 56.80 g(S=3.35) and 59.54(S=4.64) respectively. A further split was performed in node 10 taking AH as split criterion. Node 20 had eggs with AH<5.6(n=17) and node 21 had those with AH≥5.6(n=8). Predicted EW for nodes 20 and 21 came out to be 55.96 g(S=2.994) and 58.59 g(S=3.55). Nodes 3, 8, 9, 11, 20 and 21 were identified to be the terminal nodes as maximal homogeneity was attained within these nodes and no further split was possible. The highest predicted egg weight was obtained in node 3 (SW≥6.2) which was a terminal node where average predicted egg weight obtained was 63.994 g (n=13).

The relative variable importance plot for regression tree model constructed is given in Figure 2.

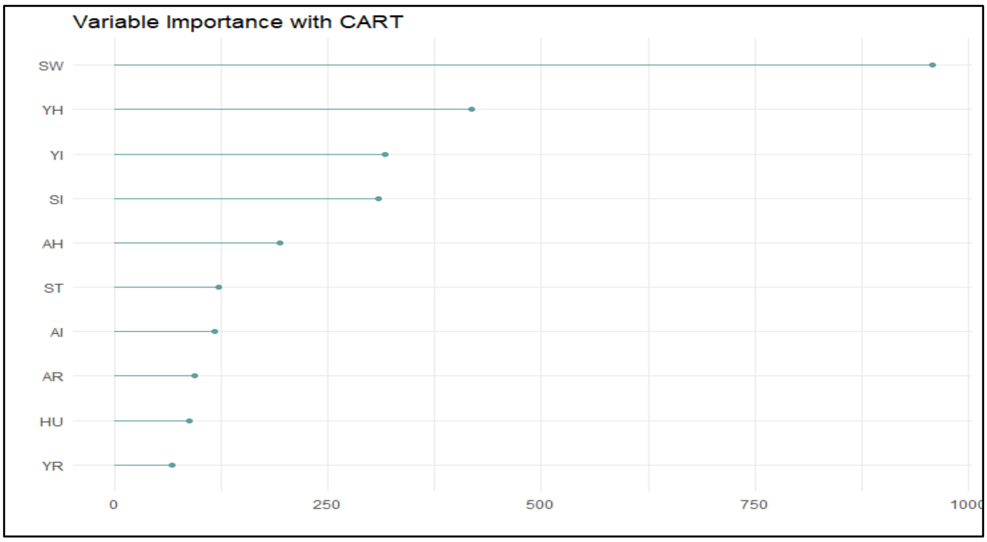

Figure 2: Variable Importance plot for CART algorithm in White leghorn

The model performance can be assessed from the predicted vs actual plot given in figure 3. It is evident that at the low end, the model overestimates, whereas at the high end, it underestimates. This indicated that the prediction model developed using CART algorithm was not much effective in predicting the egg weight in White leghorn dataset.

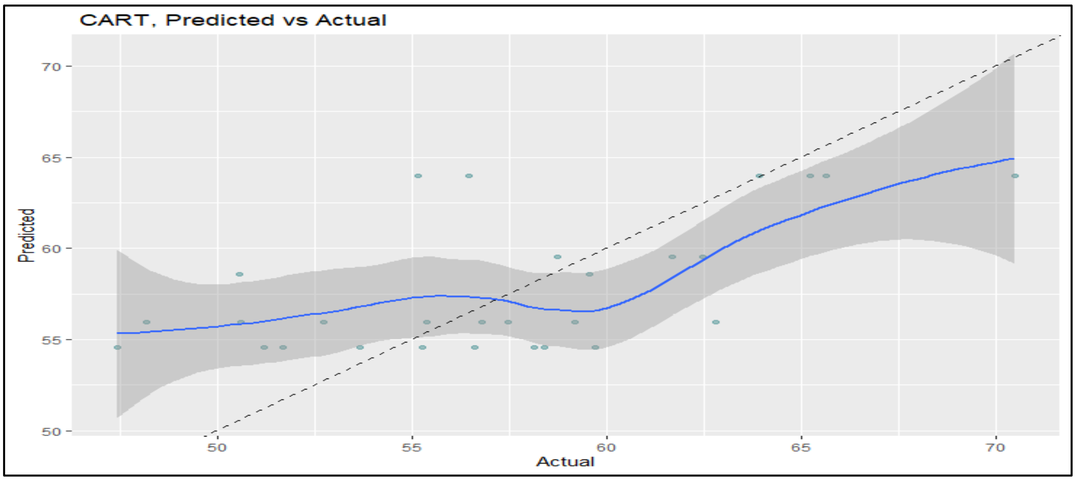

Figure 3: Predicted vs Actual Plot CART regression tree- White leghorn

Bagging and Random forests were applied to the egg quality data in White leghorn. Random Forests defaults to using p/3 variables out of total predictor variables for regression trees in a random forest. As a result, a random forest was produced on the same data set with mtry = 3. In figure 4, the test error for egg quality data in the white leghorn dataset with m=3 predictors are plotted as a function of the number of trees in the random forests. The decline in test error as number of trees increases is evident. The graph in figure 5 shows how well the random forests model performs on the test dataset. The plot obtained for variable importance using Random Forest is given in figure 6.

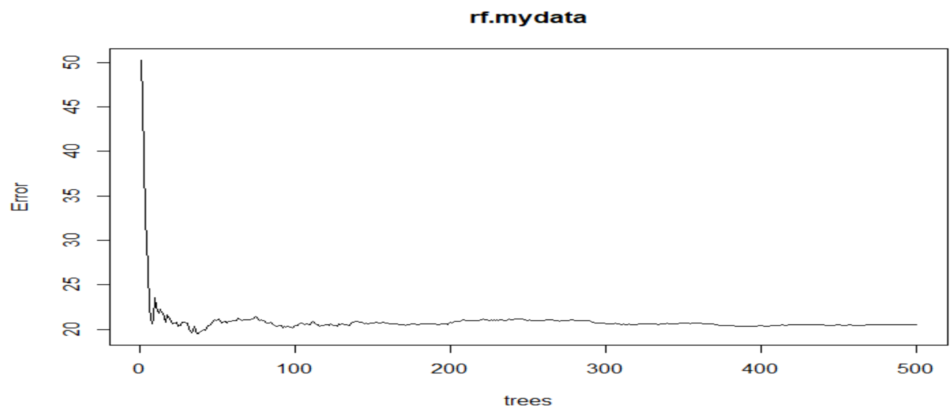

Figure 4: Test error displayed as a function of number of trees in Random Forests regression in white leghorn dataset

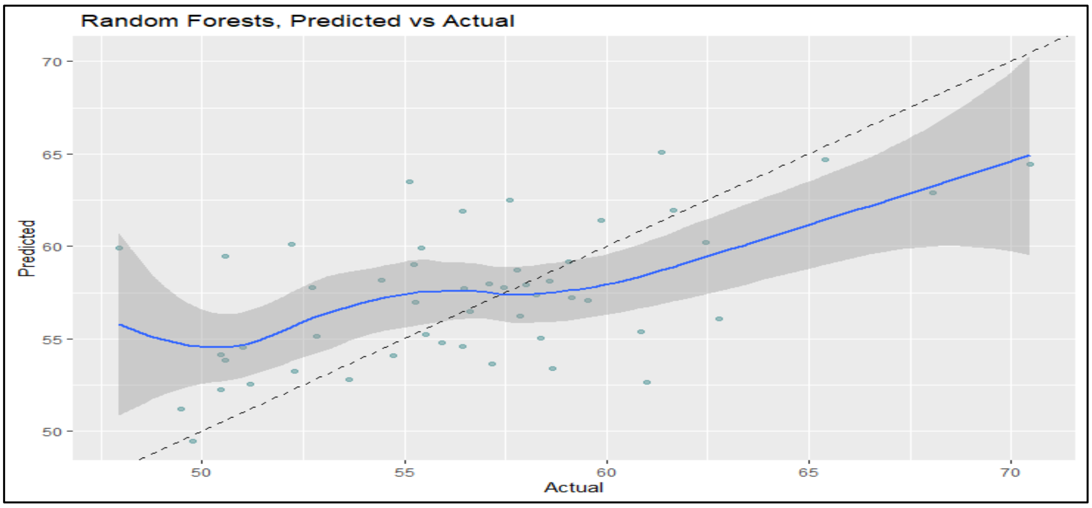

Figure 5: Actual vs Predicted plot- Random Forests in White leghorn data

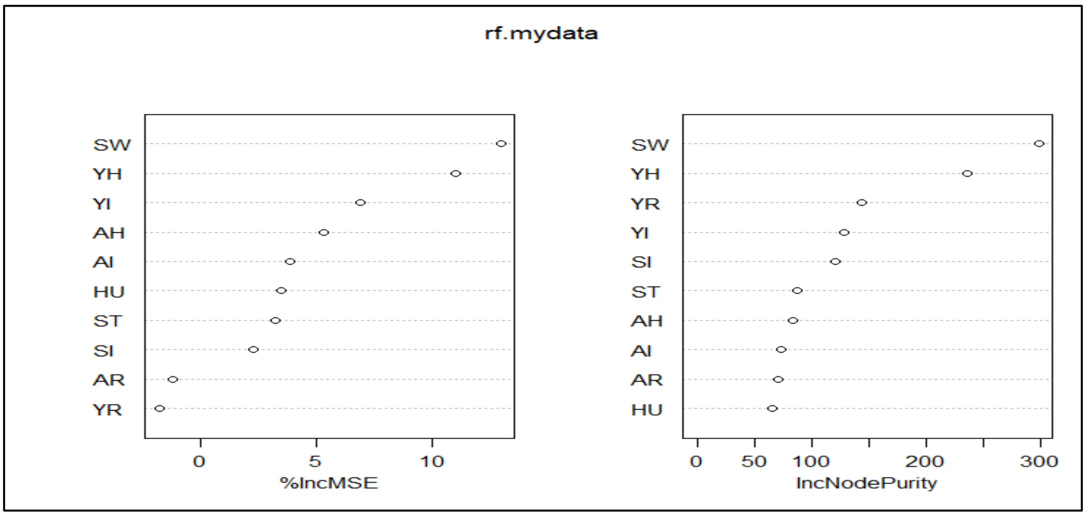

Figure 6: Variable Importance plot- Random Forests Regression in White leghorn dataset.

The comparison of predictive performance of the models employed for the present study was performed using the following model evaluation criteria.

Table 5: Goodness of fit measurements of fitted models in White leghorn dataset

Algorithm	**R^2**	**Adj. R^2**	**RMSE**	**MSE**	**MAE**
MLR	88.65	86.43	1.49	2.20	1.28
RIDGE	83.49	81.13	2.07	4.29	1.65
LASSO	31.75	29.67	4.21	17.75	3.48
ELASTIC NET	87.87	86.5	1.78	3.15	1.43
CART	33.79	29.45	4.35	37.21	4.97
RANDOM FORESTS	66.001	65.02	3.92	15.42	4.15

The highest Adj. R^2 values were observed for elastic net (86.5), MLR (86.43) and Ridge (81.3). Though the values were comparable, considering the multicollinearity, it can be stated that the elastic net method is most suited for the White leghorn dataset. The lower performance values for the LASSO is an indication that the variable selection was not effective. The poor performance for the CART regression trees is an indication that despite, its easiness of interpretation being a visual method, the CART algorithm fails in explaining the variability in EW effectively. It can also be observed that an improvement in predictive ability is notice when random forests algorithm is utilized instead of the CART.

The average EW and most of the other quality parameter averages in White leghorn breed were comparable to the results obtained by Sumashree et al., (2019). However, the mean value for AH and YH were 4.80±0.12 mm and 14.73±0.18 mm respectively which were slightly lower to that reported by Sumashree et al., (2019). Data mining algorithms CART and CHAID were used to predict the egg weight from egg quality traits in Indigenous free-range chickens of Zambia by Liswaniso et al., 2021. This study also revealed R^2, Adj. R^2 values and lower predictive performance for CART. In the study conducted by Ciftsuren et. al., 2018 to predict the internal egg quality characteristics using the OLS, Ridge, LASSO and Elastic net methods, the goodness of fit values for the methods were similar, but considering the multicollinearity, LASSO method was identified as the optimal model.

Conclusions

On performing prediction of egg weight using the identified set of egg quality traits and evaluating the predictive performance of the applied models in terms of Adj. R^2 values, comparable, high Adj. R^2 value was obtained for the Elastic net, MLR and Ridge models. Taking into account the multicollinearity that was evident in the VIF values, the Elastic net which showed the highest Adj. R^2 was identified to be best suited predictive model for the current White leghorn dataset. This demonstrates that the Elastic Net model's shrinkage and regularization were comparatively more effective in the current prediction problem. The variable selection with LASSO method was found to have lower predictive ability in the dataset analysed. The performance of CART algorithm was poor and the predictive ability was found to improve on using random forests instead of CART. It could be concluded that the predictive ability of any model for any dataset is largely dependent on the structure of the data, nature of predictors and their inter-relations. The shell weight (SW) was identified to be the most crucial predictor of egg weight. The information on the relative importance of the various predictors can be tremendously useful in the development of quality standards for both table eggs and hatching eggs, and more research in this area is needed.

Author’s Contribution

The first author and the corresponding author are the main author of this article who had plan the work, collected data, done statistical analysis and present the results and rest authors have contributed equally. All authors read and approved the final manuscript.

Acknowledgements

The work has been done under MVSc reasech work of Animal Genetics & Breeding Department funded by the Central Agricultural University (CAU), Imphal, Manipur and DBT Star College Project, Government of India. The authors express their gratitudes to Dean, CVSc. & A.H., Aizawl Mizoram, for providing all the necessary support to carry out the study.

Conflict of Interests

There is no conflict of interest.

Publisher Disclaimer

IJLR remains neutral concerning jurisdictional claims in published institutional affiliation.

References

1. Celik, S., Sogut, B., Sengul, T., Eyduran, E. and Sengul, A. Y. (2016). Usability of CART algorithm for determining egg quality characteristics influencing fertility in the eggs of Japanese quail. *Revista Brasileira de Zootecnia*, **45**(11): 645-649.
2. Ciftsuren, M.N. and Akkol, S. (2018). Prediction of internal egg quality characteristics and variable selection using regularization methods: ridge, LASSO and elastic net. *Archives Animal Breeding*, **61**(3): 279-284.
3. Faridi, A., France, J. and Golian, A. (2013). Neural network models for predicting early egg weight in broiler breeder hens. *Journal of Applied Poultry Research*, **22**(1):1-8.
4. Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods and Algorithms. *John Wiley & Sons, Inc., USA.*

5. Khan, M.A., Tariq, M.M., Eyduran, E., Tatliyer, A., Rafeeq, M., Abbas, F., Rashid, N., Awan, M.A. and Javed, K. (2014). Estimating body weight from several body measurements in Harnai Sheep without multicollinearity problem. *Journal of Animal and Plant Sciences*, **24**(1): 120-126.
6. Khan, M.K.I., Khatun, M.J. and Kibria, A.K.M.G. (2004). Study the quality of eggs of different genotypes of chickens under semi-scavenging system at Bangladesh. *Pakistan Journal of Biological Sciences*, **7**(12): 2163-2166.
7. Lahmann, N.A. and Kottner, J. (2011). Relation between pressure, friction and pressure ulcer categories: a secondary data analysis of hospital patients using CHAID methods. *International Journal of Nursing Studies*, **48**(12): 1487-1494.
8. Liswaniso, S., Qin, N., Tyasi, T. L., Chimbaka, I. M., Sun, X., and Xu, R. (2021). Use of Data Mining Algorithms Chaid and Cart in Predicting Egg Weight from Egg Quality Traits of Indigenous Free-Range Chickens in Zambia. *Advances in Animal and Veterinary Sciences*, **9**(2): 215-220.
9. McDonald, G.C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, **1**(1): 93-100.
10. Okoro, V.M.O., Ravhuhali, K.E., Mapholi, T.H., Mbajiorgu, E.F. and Mbajiorgu, C.A. (2017). Comparison of commercial and locally developed layers' performance and egg size prediction using regression tree method. *Journal of Apllied Poultry Research*, **26**(4): 476-484.
11. Ranstam, J. and Cook, J.A. (2018). LASSO regression. *British Journal of Surgery*, **105**(10): 1348-1348.
12. Sumashree, N., Sharanagouda, H., Udaykumar, N., Ramappa, K.T. and Naik, N. (2019). Study on physical properties of exotic and indigenous breed eggs. ***International Journal of Current Microbiology and Applied Sciences*, 8**(9):327-334.
13. Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **67**(2):301-320.
14. Salgado, P.R., Vazquez, C., Sesma, B., Mendoza, N.P., Marin, H.A., Hernandez, E.W and Ortega, B.J. (2021). Prediction of Egg Weight from External Egg Traits of Guinea Fowl Using Multiple Linear Regression and Regression Tree Methods. *Brazilian Journal of Poultry Science*, **23**(3): 1-6.

****