



# OPEN Use of data mining algorithms in prediction of eggshell thickness from egg quality traits of Potchefstroom Koekoek layers

Kagisho Madikadike Molabe<sup>1</sup>, Thobela Louis Tyasi<sup>1✉</sup> & Vusi Gordon Mbazima<sup>2</sup>

Egg quality is affected by lot of factors. Study was conducted to compare performance of data mining algorithms; Classification and regression tree (CART), Chi-square automatic interaction detection (CHAID), Exhaustive chi-square automatic interaction detection (Ex-CHAID) and Multivariate adaptive regression spline (MARS) in prediction of Potchefstroom Koekoek's eggshell thickness from egg quality traits. 350 eggs were collected at 31st to 39th week to examine the egg quality traits. MARS with  $R^2(0.86)$  revealed yolk ratio, shell weight, egg shape index, yolk ratio, shell ratio, albumen weight and albumen ratio as explanatory variables predicting eggshell thickness. CART with  $R^2(0.37)$ , yolk/albumen ratio was noted to be influential predictor of eggshell thickness. CHAID and Ex-CHAID ( $R^2=0.35$ ) discovered egg weight as the best predictor of eggshell thickness. MARS with  $R^2(0.86)$  revealed yolk ratio, shell weight, egg shape index, yolk ratio, shell ratio, albumen weight and albumen ratio as explanatory variables predicting eggshell thickness. MARS had high  $r(0.925)$ ,  $R^2(0.856)$  and lower RMSE (0.129) and AIC (-975.331) compared to CHAID, Ex-CHAID and CART leading MARS to be the best data mining algorithm when predicting the eggshell thickness using egg quality traits.

**Keywords** Eggshell thickness, CART, CHAID, MARS, Ex-CHAID

Poultry industry as one of agricultural sector contributes to the economy through provision of food such as meat and eggs contributing 60% of protein to human's diet<sup>1,2</sup>. Indigenous chickens due to their low maintenance, disease resistance and adaptability to harsh climatic conditions, their production increase daily<sup>3</sup>. As explained by Tyasi et al.<sup>4</sup>, Potchefstroom Koekoek chicken breed is an indigenous chicken derived from crossing White Leghorn, the Black Australorp, and the Barred Plymouth Rock and it's characterised by black and white colour patterns. Potchefstroom Koekoek chicken breed is a prolific brown shelled egg layer with average weight of 55.7 g, brown egg shell results from syntheses of protoporphyrin IX pigment in the shell gland which is deposited onto all shell layers including the shell membranes, but most the calcareous shell and in the cuticle<sup>5</sup>. Egg quality is affected by lot of factors including genes and have impact on the acceptability by consumers and recently, the genes that are involved in pigment synthesis have been identified, but the genetic control of synthesis and deposition of brown pigment is not fully understood<sup>6</sup>. Literature on metabolic pathway of pigments in the shell gland and the genes which overexpression can significantly increase the intensity shell pigment and improve shell quality. Egg quality traits are constituents of the egg and are divided into internal, focusing on the content of the egg and external, focusing on the eggshell egg quality traits<sup>4</sup>. Eggshell as the closure and the protector of the internal content of the egg plays a major role in protecting the embryo from external forces and enabling development<sup>7</sup>. However, Potchefstroom Koekoek layers experience lot of damage and loss resulting from cracked eggshell due to diseases, infections, inadequate nutrition, expression of genes and environmental factors<sup>8</sup>. Egg quality has long been subjected to genetic selection, and considerable improvement of eggshell quality has been reported using linear and multiple regression<sup>9</sup>. Recently the approach of data mining algorithm gained exposure in the poultry industry where they are utilised to predict traits of economic importance such as egg weight<sup>10</sup>. In making decisions for selection to improve desired traits, breeders deal with a lot of data consisting of a lot of factors which in most cases renders the data multidimensional.

Data mining algorithms are more considered as they possess in multi-collinearity<sup>11</sup>. Additionally, as documented by Kantardzic<sup>12</sup>, these algorithms are computer-based method to discover information from data.

<sup>1</sup>Department of Agricultural Economics and Animal Production, University of Limpopo, Private Bag X1106, Sovenga 0727, South Africa. <sup>2</sup>Department of Biochemistry, Microbiology & Biotechnology, University of Limpopo, Private Bag X1106, Sovenga, Limpopo 0727, South Africa. ✉email: louis.tyasi@ul.ac.za

To the best of authors' knowledge, importance of the data mining algorithms in poultry industry had been poorly utilised hence the aim of the current study was to compare performance of data mining algorithms; Classification and regression tree (CART), Chi-square automatic interaction detection (CHAID), Exhaustive chi-square automatic interaction detection (Ex-CHAID) and Multivariate adaptive regression spline (MARS) in prediction of Potchefstroom Koekoek eggshell thickness from egg quality traits; Albumen weight (AW), Albumen ratio (AR), Yolk weight (YW), Yolk ratio (YR), Yolk/albumen ratio (YAR), egg weight (EW), egg length (EL), egg width (EWD), egg volume (EV) and egg shape index (ESI), shell weight (SW), shell ratio (SR), shell strength (SS), shell surface area (SSA), unit surface shell weight (USSW) and eggshell thickness (STA).

Reports from the study might help Potchefstroom Koekoek layers in improving the eggshell through selection of accurate data mining algorithm and selection of best traits in predicting eggshell thickness for improvement during breeding. This study will also increase the employability of data mining algorithms in the poultry industry.

## Materials and methods

### Study area

The research was conducted at the University of Limpopo experimental farm, located in Mankweng village within Capricorn District of Limpopo Province, South Africa. The farm lies at latitude of 27.55 °South and longitude 24.77 °East and experiences semi-arid climatic conditions<sup>13</sup>.

### Experimental animal, management, design

A total of one hundred ( $n=100$ ) point of lay Potchefstroom Koekoek chicken layers were bought from local market were used as experimental animals. Animals were raised under intensive management system, they were housed in cages (medium-tier cage system with 3 hens in each cage of  $30 \times 30 \times 25.5$  cm) in a house that was cleaned with water, disinfectant and left for 14 days to destroy any remaining organisms. Water and feeds (layers' mash containing 16% crude protein and 11.97 MJ/kg/DM) and water were always provided. Standard management protocol was followed as suggested by Manyelo et al.<sup>13</sup>. Complete randomise design was applied as the research design.

### Egg collection

A total of three hundred and fifth ( $n=350$ ) eggs were randomly collected from week 31st to 39th, numbered correctly and taken to laboratory for measurements of egg characteristics.

### Collection of internal egg quality traits

Albumen weight (g), Albumen ratio (%), Yolk weight(g), Yolk ratio (%), Yolk/albumen ratio (%) were measured as internal egg quality trait. The yolk and albumen were allowed to pass through a whole created on the eggs and they were separated using egg yolk separator. The albumen weight was determined by subtracting yolk weight (measured using electronic scale of crate scale 300 kg x 01 kg of Tronic services, South Africa) from shell weight. Other internal egg quality traits were determined following Eqs. (1, 2 and 3) by Ashraf et al.<sup>14</sup>.

$$\text{Albumen ratio (\%)} = \frac{\text{albumen weight}}{\text{egg weight}} \times 100 \quad (1)$$

$$\text{Yolk ratio (\%)} = \frac{\text{yolk weight}}{\text{egg weight}} \times 100 \quad (2)$$

$$\text{Yolk/Albumen (\%)} = \frac{\text{yolk weight}}{\text{albumen weight}} \times 100 \quad (3)$$

### Collection of external egg quality traits

The external egg quality traits collected were egg weight (g), egg length (mm), egg width (mm), egg volume ( $\text{cm}^3$ ) and egg shape index (%). Egg weight was weighed using balanced weighing scale (crate scale 300kgx01kg) of Tronic services, South Africa. Egg length (mm) and egg width (mm) were measured using vernier caliper. The shape index was determined using Eq. (4) as described by Ashraf et al.<sup>14</sup> whereas egg volume ( $\text{cm}^3$ ) was determined following Eq. (5) by Kgatalala et al.<sup>15</sup>.

$$\text{Shape index (\%)} = \frac{\text{egg width}}{\text{egg length}} \times 100 \quad (4)$$

$$\text{Egg volume} = [0.6057 - (0.018 - \text{Egg width} * \text{Egg length} * \text{egg width}^2)]. \quad (5)$$

### Collection of eggshell quality traits

The shell weight ( $\text{g/cm}^2$ ), shell ratio (%), shell strength (N), shell surface area ( $\text{cm}^3$ ), unit surface shell weight ( $\text{g/cm}^2$ ) and eggshell thickness (mm) were measures. Shell weight was measured using electronic scale of scale 300 kg x 01 kg from Tronic services, South Africa. Three regions (sharp, equatorial, and blunts) of the eggshell were measured using micrometer gauge (QCT shell thickness micrometer, TSS, England) after exposure to air for 72 h to get eggshell thickness, the average was considered. Eggshell strength (N) was measured using QCT shell strength tester. Other shell quality trait was measured following Eq. (6) by Ashraf et al.<sup>14</sup>.

$$\text{Shell ratio (\%)} = \frac{\text{shell weight}}{\text{egg weight}} \times 100 \quad (6)$$

### Classification and regression tree (CART) data mining algorithm

The CART is a decision tree which split data into uniform subgroups with regards to the reliant variable<sup>16</sup>. Decision tree is constructed partitioning a node 0 (parent note) which contains all data set into child nodes ensuring minimum error variance using cross-validation training and test sets<sup>17</sup>. According to<sup>18</sup> documentation, CART algorithm, prune to avoid unnecessary nodes in the decision tree diagram and the technique aims to have a terminal node to increase proportion of differences between nodes. CART as an understandable model to predict an event on both ordinal and nominal scale<sup>16</sup>. The goal of this technique is to have terminal nodes to increase the proportion of differences between nodes<sup>18</sup>.

### Chi-squared automatic interaction detection (CHAID) data mining algorithm

The CHAID divide population into categories such that the variation in a dependent variable within groups is minimized and among groups is maximized<sup>10</sup>. It's a prediction model that defines the outcome in each dependent variable, moreover ordinal, continuous and nominal data can be used<sup>19</sup>.

### Exhaustive CHAID (Ex-CHAID) data mining algorithm

The Ex-CHAID) is the extension of CHAID algorithm and aim is to maximize the variance between nodes, depending upon their tree depth<sup>20</sup>. It expresses the same splitting and stopping steps as CHAID however it has got the merging step more exhaustive than CHAID<sup>21</sup>.

### Multivariate Adaptive Regression Splines (MARS) data mining algorithm

The MARS data mining algorithm as a form of regression analysis developed in 1991 by Friedman<sup>22</sup>. MARS is a non-parametric regression procedure which can produce linear models that are nonlinearities and interactions between variables<sup>11,23</sup>. Its advantage is that it can overcomes the multicollinearity problems<sup>17</sup>. The MARS algorithm was conducted as explained by<sup>24</sup>, and its prediction equation can be written as follows:

$$f(x) = \beta_0 + \sum_{m=1}^m \beta_m \lambda_m(x) \quad (7)$$

where  $f(x)$  is the expected response,  $\beta_0$  and  $\beta_m$  are parameters that are calculated to give the best data fit, and  $m$  is the number of BF's in the model. In the MARS model, the basis function composed of be a single univariable spline function or a combination of more than one spline function for diverse predictor inputs. The spline BF,  $\lambda_m(x)$ , is defined as:

$$\lambda_m(x) = \prod_{k=1}^{k_m} [s_{km}(X_{v(k,m)} - t_{k,m})] \quad (8)$$

where  $t_{k,m}$  denotes the knot location;  $s_{km}$  denotes the right/left regions of the corresponding step function, taking either 1 or -1;  $v(k, m)$  denotes the predictor variable's label; and  $k_m$  is the number of knots. The pruning process was used to remove the basic functions that had a low contribution to the model fitting performance following the generalised cross-validation error (GCV)<sup>24</sup>:

$$GCV(\lambda) = \frac{\sum_{i=1}^n (y_i - \hat{y}_{ip})^2}{(1 - \frac{M(\lambda)}{n})^2} \quad (9)$$

where  $n$  represents the number of training cases,  $y_i$  shows the observed value of the responsible variable,  $\hat{y}_{ip}$  as the estimated value of the response variable and  $M(\lambda)$  represents the penalty function for the complex of the model with  $\lambda$  terms.

### Performance assessment of used data mining algorithms

The goodness of fit test was used to compare the prediction performance of CART, CHAID, Ex-CHAID and MARS as suggested by<sup>11</sup>.

Pearson's correlation coefficient ( $r$ ):

$$r = \frac{cov(y_i y_{ip})}{S_{y_i} S_{y_{ip}}} \quad (10)$$

Coefficient of Determination ( $R^2$ ):

$$Rsqr = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

Adjusted Coefficient of Determination (Adj. $R^2$ ):

$$Adj.R^2 = 1 - \frac{\frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

Akaike Information Criteria (AIC):

$$AIC = n \ln \left[ \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n} \right] + 2k \quad (13)$$

Root-mean-square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

### Statistical analysis

The Statistical Package for Social Sciences (IBM SPSS, 2022) version 29.0 software was used for data analysis at significance level of 5%<sup>25</sup>. The data mining algorithms such as CART, CHAID, Ex-CHAID and MARS were used to achieve the objective of this study.

## Results

### CART data mining algorithm

Decision tree (CART) containing detailed information on the independent traits significantly affecting eggshell thickness is displayed in Fig. 1. Parent (root) node showed an average mean of shell thickness as 0.540 with standard deviation of 0.331 with total number of 350 eggs included. The root node was split into subgroups, node 1 with average mean of 0.427 and standard deviation of 0.241, node 2 with average and standard deviation of 0.769 and 0.370 respectively according to the eggshell thickness with YAR and EW as the independent traits. Node 2 was further divided to node 3 (mean of 0.658 and standard deviation of 0.331) and node 4 having a mean of 1.174 and standard deviation of 0.159 with egg weight as the independent trait. The thickest eggshell thickness mean was noted on node 4 (1.174) with YAR > 70.595 and EW > 53.260. Lowest eggshell mean was observed on node 1 (0.427) with YAR ≤ 70.595.

### CHAID data mining algorithm

Regression tree constructed by CHAID decision tree is displayed in Fig. 2. The overall average in the parent node with eggshell thickness as the dependent traits shown to be 0.540 with standard deviation of 0.331. Total of 350 eggs included in the study were group according to eggshell thickness and divided into subgroups called nodes. The parent node (node 0) was split to 8 nodes with EW as the independent trait.

The highest average on eggshell thickness was obtained on node 6 (1.018) with EW between 52.790 and 53.630 and the lowest average on eggshell thinness was noted on node 2 (0.313) with EW between 45.510 and 46.590.

### Ex-CHAID data mining algorithm

In the Exhaustive CHAID (Ex-CHAID) in Fig. 3 noted that, an average mean (0.54) and standard deviation (0.33) were observed on node 0 with STA as the dependent trait and 350 eggs were present in the top selection generated. The node 0 was subdivide into 7 nodes based on eggshell thickness with EW as the independent trait. The lowest average on eggshell thickness was observed on node 2 (0.313) with EW between 45.510 and 46.590 and the highest average on eggshell thickness was noted on node 5 (1.018) with EW between 52.790 and 53.630.

### MARS data mining algorithm

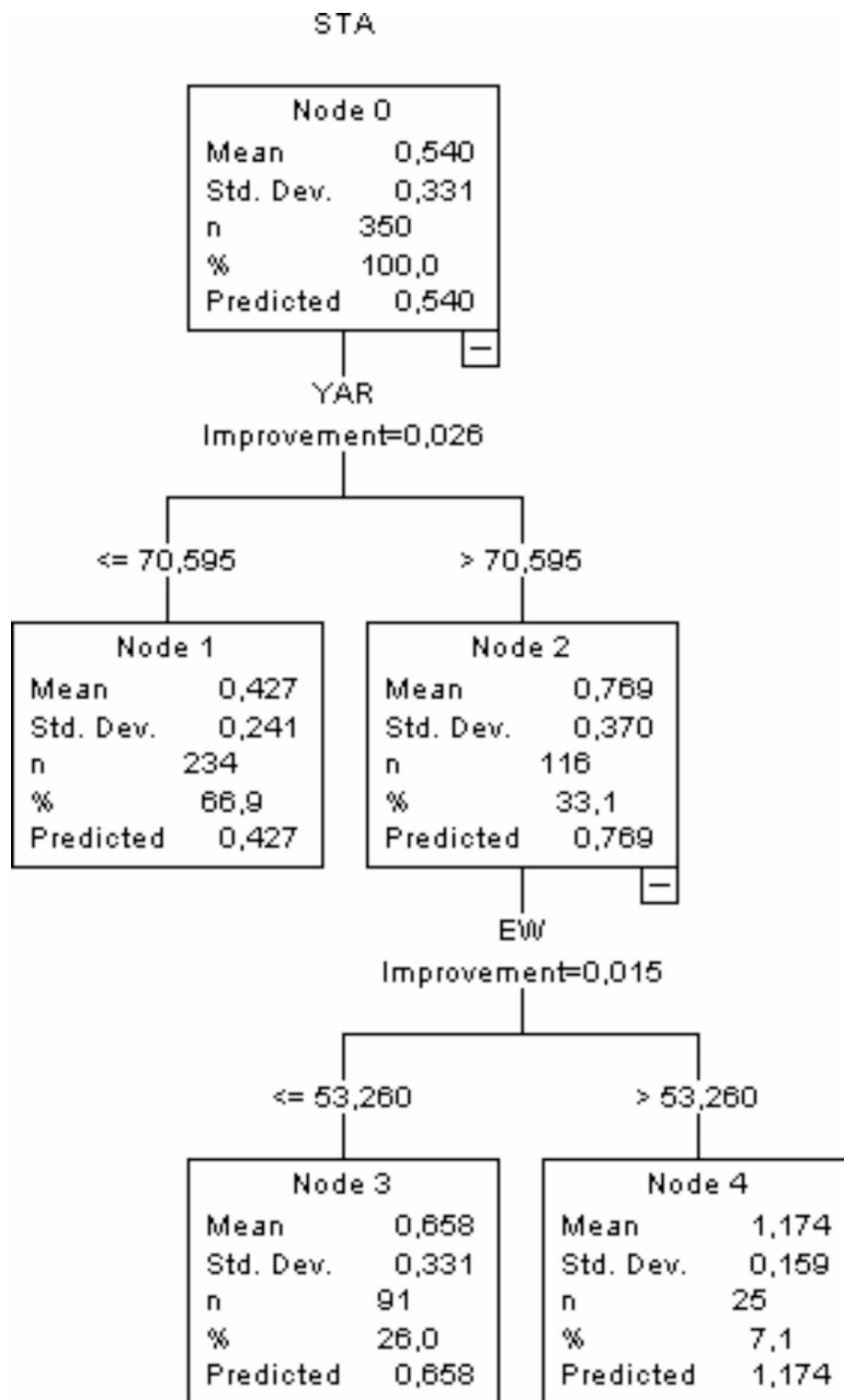
Multivariate adaptive regression splines (MARS) algorithm reports showed that YR, SW, ESI, YR, SR, AW and AR were involved in the model (Table 1). Results showed eighteen basic functions from MARS model, three as single order term variables and fifteen as orders of interactions with an intercept of 0.307. MARS displayed positive and negative coefficient showing impact of egg quality traits on eggshell thickness. Positive influence on eggshell thickness was noted when YW < 15.53 with model coefficient of 0.372. Moreover, the model showed the effect of egg quality traits interactions on STA. Briefly, the influence on STA was 0.020 when YW > 15.53 and YR > 37.14, 0.033 when YW > 15.33 and SR > 36.3092, 0.017 when ESI < 71.8799, YW > 15.53 and YR < 37.14, 0.021 when ESI > 71.48, YW > 15.53, AR > 53.58, 0.119 when YW > 15.53, YR > 37.14, SW < 5.47 and lastly 0.016 when YW > 15.53, YR < 37.14, SR < 35.48.

### Data mining algorithms performance

The predictive and statistical performance using several goodness of fit criteria are presented in Table 2. MARS was estimated to have high  $r$ ,  $R^2$ , Adjusted  $R^2$  and lower RMSE and AIC as compared to CHAID, Ex-CHAID and CART, as such lead MARS to be the best accurate data mining algorithm when predicting the eggshell thickness using egg quality traits.

## Discussion

Data mining algorithms provide reliable information on relation between traits as they cater for multicollinearity problem among traits than regression techniques<sup>26</sup>. Hence the current study compares performance of data mining algorithms including CART, CHAID, Ex-CHAID and MARS on prediction of Potchefstroom Koekoek's eggshell thickness from egg quality traits. The output on CART data mining algorithm with  $R^2$  of 0.37, yolk/albumen ratio was noted to be influential predictor of eggshell thickness. Finding further showed that both CHAID and Ex-CHAID data mining algorithms ( $R^2 = 0.35$ ) discovered egg weight as the best predictor of eggshell thickness. While MARS with  $R^2$  of 0.86 revealed yolk ratio, shell weight, egg shape index, yolk ratio, shell ratio, albumen weight and albumen ratio as explanatory variables predicting eggshell thickness. There is



**Fig. 1.** CART data mining algorithm.

no literature on use of data mining algorithm in prediction of eggshell from egg quality traits, however data mining algorithm had been applied and explored in the poultry industry in prediction of egg weight. Due to lack of literature on the prediction of eggshell thickness using data mining algorithms, current findings were discussed based on comparing current study results using studies conducted on predicting egg weight from egg quality traits using data mining algorithm. Study by<sup>19</sup> compared CHAID and CART in predicting egg weight from egg quality traits on indigenous free-range chickens and found that CHAID algorithm had the highest  $R^2$ (0.82) and it was recommended for prediction of egg weight. Study conducted by<sup>10</sup> on predictive performance

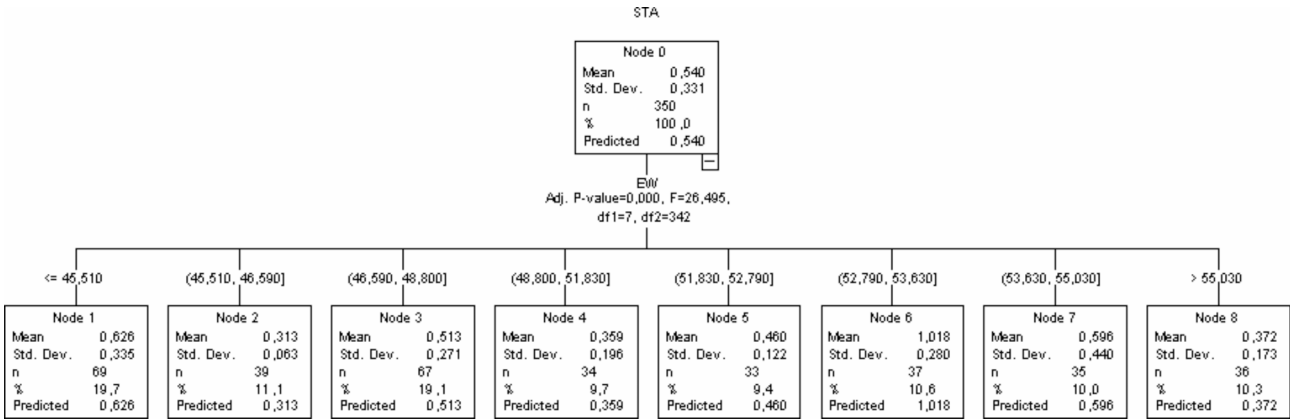


Fig. 2. CHAID data mining algorithm.

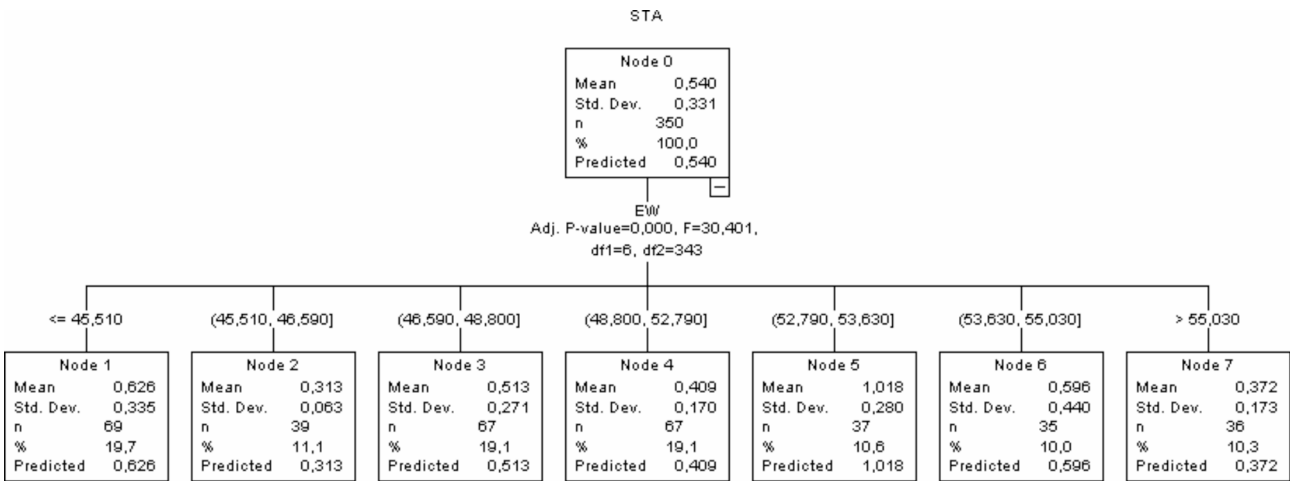


Fig. 3. Ex-CHAID data mining algorithm.

| BF        | Equation                                   | Coefficient |
|-----------|--|-------------|
| Intercept |  | 0.307       |
| BF1       | h(YW-15.53)                                | 0.372       |
| BF2       | h(6.55-SW)                                 | 0.147       |
| BF3       | h(SW-6.55)                                 | 0.137       |
| BF4       | h(71.48-ESI) * h(YW-15.53)                 | -0.021      |
| BF5       | h(YW-15.53) * h(YR-37.14)                  | 0.020       |
| BF6       | h(YW-15.53) * h(37.14-YR)                  | -0.059      |
| BF7       | h(YW-15.53) * h(SR-36.3092)                | 0.033       |
| BF8       | h(YW-15.53) * h(36.3092-SR)                | -0.050      |
| BF9       | h(71.8799-ESI) * h(YW-15.53) * h(37.14-YR) | 0.017       |
| BF10      | h(ESI-71.8799) * h(YW-15.53) * h(37.14-YR) | -0.011      |
| BF11      | h(ESI-71.48) * h(YW-15.53) * h(AR-53.58)   | 0.021       |
| BF12      | h(ESI-71.48) * h(YW-15.53) * h(53.58-AR)   | -0.004      |
| BF13      | h(YW-15.53) * h(AW-27.78) * h(37.14-YR)    | -0.007      |
| BF14      | h(YW-15.53) * h(27.78-AW) * h(37.14-YR)    | -0.072      |
| BF15      | h(YW-15.53) * h(YR-37.14) * h(SW-5.47)     | -0.044      |
| BF15      | h(YW-15.53) * h(YR-37.14) * h(5.47-SW)     | 0.119       |
| BF17      | h(YW-15.53) * h(37.14-YR) * h(35.48-SR)    | 0.016       |

Table 1. MARS data mining algorithm.



| Goodness of fit         | CHAID    | Ex-CHAID | CART     | MARS     | Decision          |
|-------------------------|----------|----------|----------|----------|-------------------|
| r                       | 0.593    | 0.589    | 0.611    | 0.925    | Greater is better |
| Adjusted R <sup>2</sup> | 0.350    | 0.345    | 0.372    | 0.845    | Greater is better |
| R <sup>2</sup>          | 0.352    | 0.347    | 0.374    | 0.856    | Greater is better |
| RMSE                    | 0.267    | 0.268    | 0.2625   | 0.129    | Smaller is better |
| AIC                     | −923.889 | −921.488 | −936.059 | −975.331 | Smaller is better |

**Table 2.** Predictive performance of data mining algorithms.

of some data mining algorithms (CHAID, exhaustive CHAID and CART) implemented in the estimation of egg weight in quail, exhaustive CHAID model was best in predicting egg weight with  $R^2$  of 85.86. On the study that was conducted on relationship among egg quality traits in Japanese quails and prediction of egg weight and color using data mining algorithms, MARS performed better than CART in estimating egg weight in quails<sup>27</sup>. Regression techniques has been used in predicting eggshell thickness from egg quality traits. Study conducted by<sup>28</sup> on Japanese Quails reported egg length as a desirable trait for accurate prediction of eggshell thickness using linear regression model. Nordstrom et al.<sup>29</sup>highlighted on study conducted on Single Comb White laying hens that shell thickness, is more precisely estimated from egg specific gravity when egg weight is also included as the second independent variable on multiple linear regression model. Study on Japanese Quails and documented that eggshell thickness may be predicted using egg width<sup>30</sup>. Current study outcomes suggests that eggshell thickness of Potchefstroom Koekoek may be predicted from egg quality traits using data mining algorithms. The study showed that MARS ( $R^2 = 0.86$ ) was satisfactory in predicting eggshell thickness of Potchefstroom Koekoek using egg quality traits. From MARS model, yolk ratio, shell weight, egg shape index, yolk ratio, shell ratio, albumen weight and albumen ratio showed contribution to eggshell thickness. Potchefstroom Koekoek farmers might use the findings from the current study to improve eggshell thickness of their chicken by selecting traits (Yolk ratio, shell weight, egg shape index, yolk ratio, shell ratio, albumen weight and albumen) that showed influence on eggshell thickness. Several studies were conducted on prediction of egg weight using data mining algorithm while prediction on eggshell thickness is neglected, we recommend that more studies should be conducted on the prediction of eggshell thickness using data mining algorithm on different chicken breeds.

Conclusions

In the current study, data mining algorithm may be used to predict eggshell thickness using egg quality traits. MARS was estimated to have high r (0.925),  $R^2$ (0.856), Adjusted  $R^2$  (0.845) and lower RMSE (0.129) and AIC (−975.331) as compared to CHAID, Ex-CHAID and CART. The study concludes that MARS data mining algorithm was more informative in the predictive accuracy of eggshell, as such lead MARS to be the best accurate data mining algorithm when predicting the eggshell thickness using egg quality traits.

Recommendation

Therefore, it is recommended that chicken breeders may consider applying MARS algorithm for the prediction of eggshell thickness. More studies need to be done on predicting eggshell thickness in the poultry industry.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 18 September 2024; Accepted: 10 January 2025  
Published online: 11 January 2025

References

1. Sun, C., Liu, J., Yang, N. & Xu, G. Egg quality & egg albumen property of domestic chicken, duck, goose, Turkey, quail, & pigeon. *Poult. Sci.* **98**, 4516–4521 (2019).
2. Ochieng et al. Mycotoxins in poultry feed & feed ingredients from Sub-saharan Africa & their impact on the production of broiler & layer chickens: a review. *Toxins* **13**(9), 633 (2021).
3. Manyelo, T. G., Selaledi, L., Hassan, Z. M. & Mabelebele, M. Local chicken breeds of Africa: their description, uses & conservation methods. *Animals* **10**(12), 2257 (2020).
4. Tyasi, T. L., Lebogang, L. & Hloko, V. R. Comparative study of egg quality traits between Potchefstroom Koekoek & Hy-line silver brown layers. *Bulg. J. Agric. Sci.* **28**(1), 145–150 (2022).
5. FAO. Family Poultry Development—Issues, Opportunities, & Constraints; Animal Production & Health Working Paper No. 12; Food & Agriculture Organization of the United Nations: Rome, Italy (2014).
6. Okon, B., Ibom, L. A., Dauda, A. & Ebegbulem, V. N. Egg quality traits, phenotypic correlations, egg & yolk weights prediction using external & internal egg quality traits of Japanese quails reared in Calabar, Nigeria. *Int. J. Mol. Biol.* **5**(1), 21–26 (2020).
7. Kibala, L., Rozempolska-Rucinska, I., Kasperek, K., Zieba, G. & Lukaszewicz, M. Eggshell qualities as indicative of eggshell strength for layer selection. *Braz J. Poult. Sci.* **20**, 99–102 (2018).
8. Ledvinka, Z., Zita, L. & Klesalová, L. Egg quality and some factors influencing it: a review. *Sci. Agric. Bohem.* **43**, 46–52 (2012).
9. Liu et al. Estimation of genetic parameters of eggshell translucency & production traits in different genotypes of laying hens. *Poult. Sci.* **102**(5), 102616 (2023).
10. Celik et al. Estimation of egg weight from some external & internal quality characteristics in quail by using various data mining algorithms. *Indian J. Anim. Sci.* **87**(12), 1524–1530 (2017).

11. Phaladi, A. M., Tyasi, T. L., Tada, O. & Mogashoa, S. Use of data mining algorithms to predict body weight of dorper sheep breed in two selected villages of Lepelle-Nkumpi local municipality in Limpopo province, South Africa. *Smart Agr Technol.* **8**, 00479 (2024).
12. Kantardzic, M. Data mining: Concepts, models, methods, & algorithms (John Wiley & Sons, 2011) <https://doi.org/10.1002/9781118029145>.
13. Manyelo, T. G., Sebola, N. A., Ng'ambi, J. W., Weeks, W. & Mabelebele, M. The influence of different amaranth leaf meal inclusion levels on performance, blood profiles, and gut organ characteristics of Ross 308 broiler chickens. *Front. vet. sci.* **9**, 869149 (2022).
14. Ashraf, S., Ashraf, M., Akram, M., Malik, S. & Altaf, M. Influence of egg weight on egg quality parameters & growth traits in Ring necked pheasants in captivity. *J. Anim. Plant. Sci.* **26**, 331–338 (2016).
15. Kgwatlala, P. M., Bolebano, L. & Nsoso, S. J. Egg quality characteristics of different varieties of domesticated helmeted guinea fowl (Numida meleagris). *Int. J. Poult. Sci.* **12**(4), 245–250 (2013).
16. Olfaz, M., Tirink, C. & Önder, H. Use of CART & CHAID algorithms in Karayaka Sheep Breeding. *Kafkas Univ. Vet. Fak Derg.* **25**(1), 105–110 (2019).
17. Mathapo, M. C., Mugwabana, T. J. & Tyasi, T. L. Prediction of body weight from morphological traits of South African non-descript indigenous goats of Lepelle-Nkumbi Local Municipality using different data mining algorithm. *Trop. Anim. Health Pro.* **54**(2), 102 (2022).
18. Koc, Y. Application of regression tree method for different data from animal science. MSc Thesis, Iğdir University **58** (2016).
19. Liswaniso, S., Qin, N., Tyasi, T. & Chimbaka, I. Use of data mining algorithms chaid & cart in predicting egg weight from egg quality traits of indigenous free-range chickens in Zambia. *Adv. Anim. Vet. Sci.* **9**(2), 215–220 (2021).
20. Ali et al. Comparison of artificial neural network & decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. *Pakistan J. Zool.* **47**, 1579–1585 (2015).
21. Orhan, H., Eydurhan, E., Tatliyer, A. & Saygici, H. Prediction of egg weight from egg quality characteristics via ridge regression & regression tree methods. *Rev. Bras. De Zootec.* **45**(7), 380–385 (2016).
22. Tyasi et al. Multivariate adaptive regression splines data mining algorithm for prediction of body weight of Hy-line silver brown commercial layer chicken breed. *Adv. Anim. Vet. Sci.* **8**(8), 794–799 (2020).
23. Nieto et al. A new predictivemodel of centreline segregation in continuous cast steel slabsby using multivariate adaptive regression splines approach. *Materials* **8**, 3562–3583 (2015).
24. Sengül, T., Çelik, S. & Sengül, Ö. Use of multivariate adaptive regression splines (MARS) for predicting parameters of breast meat in quails. *J. Anim. Plant. sci.* **30**, 786–793 (2020).
25. IBM Corp. *IBM SPSS Statistics for Windows, Version 29.0.2.0 Armonk* (IBM Corp, 2022).
26. Mohammad et al. Predicting of body weight from body measurements using regression (RT) method for indigenous sheep breeds in Balochistan. *J. Anim. Plant. Sci.* **22**(1), 20–24 (2012).
27. Çelik, Ş., Eydurhan, E., Şengül, A. Y. & Şengül, T. Relationship among egg quality traits in Japanese quails and prediction of egg weight and color using data mining algorithms. *Trop. Anim. Health Pro.* **53**(3), 382 (2021).
28. Alkan, S., Karabag, K., Galic, A. & Balcioğlu, M. S. Predicting yolk height, yolk width, albumen length, eggshell weight, egg shape index, eggshell thickness, egg surface area of Japanese quails using various eggs traits as regressors. *Int. J. Poult. Sci.* **7**(1), 85–88 (2008).
29. Nordstrom, J. O. & Ousterhout, L. E. Estimation of shell weight & shell thickness from egg specific gravity and egg weight. *Poult. Sci.* **61**(10), 1991–1995 (1982).
30. Khurshid, A., Farooq, M., Durrani, F. R., Sarbiland, K. & Chand, N. Predicting egg weight, shell weight, shell thickness and hatching chick weight of Japanese quails using various egg traits as regressors. *Int. J. Poult. Sci.* **2**, 164–167 (2003).

## Acknowledgements

We thank the University of Limpopo, Department of Agricultural Economics and Animal Production for allowing the proceedings of the study.

## Author contributions

M.K.M: wrote the main manuscript, Methodology, Validation, Data curation, Investigation. T.L.T. and V.G.M. revised the manuscript, Conceptualization, Funding acquisition, Supervision, Visualization, Validation, Conceptualization. All authors reviewed the manuscript and approved the submission.

## Funding

The study with reference (PMDS22051410688) was funded by South African National Research Foundation (NRF).

## Declarations

## Competing interests

The authors declare no competing interests.

## Statement of Animal rights

The study before commencement was approved by University of Limpopo Animal Research and Ethics Committee (ULAREC). The study followed ethical standards.

## Ethical approval

University of Limpopo Animal Research and Ethics Committee (ULAREC) approved the project with project number: AREC/55/2023: PG.

## Additional information

**Correspondence** and requests for materials should be addressed to T.L.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025