Predicting hatchability of layer breeders and identifying effects of animal related and environmental factors

I. Bouba,*,†,[1] B. Visser ,* B. Kemp,‡ T. B. Rodenburg ,†,‡ and H. van den Brand‡

*Hendrix Genetics, Boxmeer, 5831 CK, Netherlands; †Animals in Science and Society, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands; and ‡Adaptation Physiology Group, Wageningen University & Research, Wageningen, Netherlands

ABSTRACT In this study, a data driven approach was used by applying linear regression and machine learning methods to understand animal related and environmental factors affecting hatchability. Data was obtained from a parent stock and grand-parent stock hatchery, including 1,737 batches of eggs incubated in the years 2010−2018. Animal related factors taken into consideration were strain (white vs. brown strain), breeder age, and egg weight uniformity at the start of incubation, whereas environmental factors considered were length of egg storage before incubation, egg weight loss during incubation and season. Effects of these factors on hatchability were analyzed with 3 different models: a linear regression (**LR**) model, a random forest (**RF**) model and a gradient boosting machine (**GBM**) model. In part one of the study, hatchability was predicted and the performance of the models in terms of coefficient of determination ($\mathbf{R^2}$) and root mean square error (**RMSE**) was compared. The ensemble machine learning models (RF: $R^2 = 0.35$, RMSE = 8.41; GBM: $R^2 = 0.31$, RMSE = 8.67) appeared to be superior than the LR model ($R^2 = 0.27$, RMSE = 8.92) as indicated by the higher R^2 and lower RMSE. In part 2 of the study, effects of these factors on hatchability were investigated more into detail. Hatchability was affected by strain, breeder age, egg weight uniformity, length of egg storage and season, but egg weight loss didn't have a significant effect on hatchability. Additionally, four 2-way interactions (breeder age × egg weight uniformity, breeder age × length of egg storage, breeder age × strain, season × strain) were significant on hatchability. It can be concluded that hatchability of parent stock and grand-parent stock layer breeders is affected by several animal related and environmental factors, but the size of the predicted effects varies between the methods used. In this study, 3 models were used to predict hatchability and to analyze effects of animal related and environmental factors on hatchability. This opens new horizons for future studies on hatchery data by taking the advantage of applying machine learning methods, that can fit complex datasets better than LR and applying statistical analysis.

Key words: incubation, hatchability, machine learning, genetic, environment

2021 Poultry Science 100:101394
https://doi.org/10.1016/j.psj.2021.101394

INTRODUCTION

Effects of animal related and environmental factors on hatchability in layer breeder eggs are considerably less investigated than broiler breeder eggs. Precise prediction of hatchability may help the industry in proper prediction of hatchability of batches of eggs from the incubator (important for logistics) and possibilities to adjust incubation conditions to get more favorable outcomes. New technologies in the hatchery generate data about breeders, eggs, incubators, and other potentially important factors, that can be used for predicting performance of the hatchery (Klein et al., 2020). This abundance of data, produced in hatcheries, may require new modeling techniques (Ren et al., 2020). Machine learning techniques might be good alternatives to the classical techniques, such as linear regression, because they are based on pattern recognition (Coronel-Reyes et al., 2018). Hatchability depends on numerous animal related and environmental factors (King'ori, 2011). Some animal related factors, like breeder age and egg weight uniformity, and environmental factors, like egg storage duration, season, and egg weight loss during incubation have been shown to have an effect on hatchability, particularly in broiler breeders (Grochowska et al., 2019). Effects of these factors in layer breeders, including the effect of strain (brown vs. white) have hardly been investigated (Machado et al., 2020). In general, brown genetic

Received April 28, 2021.
Accepted July 20, 2021.

[1]Corresponding author: Ismalia.Bouba@hendrix-genetics.com

strain showed a higher hatchability than white genetic strain (Tona et al., 2010), which might be related to differences in egg characteristics (Narushin et al., 2016). Breeder age has been shown to affect hatchability in both broiler and layer breeders (Nasri et al., 2020a). In general, prime breeder flocks (31−42 wk of age) showed the highest hatchability, whereas younger (25−30 wk of age) and older breeder flocks (>42 wk of age) showed lower hatchability (Damaziak et al., 2021). Another aspect that is hardly investigated in layer and broiler breeders is the effect of egg weight uniformity on hatchability. Previous studies have focused mostly on body weight uniformity (Abbas et al., 2010) of the broiler breeder flock with little focus on egg weight uniformity. Furthermore, effect of egg storage duration on hatchability was significant. Short egg storage duration (<7 d) seems to increase hatchability of eggs from young breeders, probably because of albumen liquefaction with consequently better oxygen availability for the embryo. Prolonged (>7 d) storage reduced hatchability of eggs from 92 to 71% (Dymond et al., 2013), especially in eggs from older breeders. This might be related to modifications in the blastoderm or in the albumen during storage, which subsequently had an effect on hatchability negatively (Pokhrel et al., 2018; Nasri et al., 2020b). Egg weight loss is another important factor that could indirectly affect hatchability (Duman and Şekeroğlu, 2017). Season of incubation is another factor that has been suggested to affect hatchability; hatchability was higher during late summer than during spring and eggs from older breeders (61−65 wk of age) were more sensitive to season than eggs of younger (25−30 wk of age) breeders (Yassin et al., 2008). The aim of this study is 2-fold. The first aim is to analyze and predict hatchability of layer breeders, using 3 different models including, random forest (**RF**), gradient boosting machine (**GBM**) and linear regression (**LR**). The second aim is to investigate more specific effects of animal related (strain, breeder age, egg weight uniformity at the start of incubation) and environmental (length of egg storage before incubation, egg weight loss during incubation, seasons) factors on hatchability of layer breeders. These effects are studied in experimental setups not including all potential influencing factors. The approach has been described in material and methods.

MATERIALS AND METHODS

Dataset

A dataset (Hendrix Genetics, Boxmeer, the Netherlands) obtained from a parent (**PS**) and grand-parent stock (**GPS**) of the layer hatchery, including 1,737 batches of 100 to 166 number of eggs incubated during 2010 to 2018, was used. The dataset consisted of 3 groups of factors: animal related factors (parent stock or GPS generation, genetic strain, breeder farm, breeder age, egg weight, and egg weight uniformity at set), environmental factors (length of egg storage before incubation, set month (season), egg weight loss during incubation, day of incubation at transfer from setters to hatchers), and hatchery factors (setter number and hatcher number in which the eggs were incubated). During incubation, the temperature, CO_2, relative humidity, and airflow for the setters were set to 99°F, 0.4%, 66%, and 36 ft^3, respectively, while those of the hatchers were set to 98°F, 0.3%,50 %, 55 ft^3, respectively. Hatchability was calculated per batch of eggs as hatch of fertile eggs (**HOF**) as shown in Equation (1).

$$HOF = \left(\frac{CP + NC}{SE \; - IE}\right) * 100, \tag{1}$$

Where CP = Number of chicks pulled, NC = Number of chicks culled, SE = Number of set eggs and IE = Number of infertile eggs.

In the first part of this study, which aimed to predict hatchability, all factors in all 3 groups were used. In the second part of this study, some animal related and environmental factors were investigated more into depth on their effects on hatchability.

Overview of Methods

The overall flowchart of the used methods in both parts of this study is shown in Figure 1. For each part, the dataset, feature engineering process, dimensionality reduction techniques, methods used, and postprocessing strategies have been shown. The first part deals with predicting hatchability by using RF, GBM, and LR models. The second part deals with investigating effects of animal related and environmental factors on hatchability by applying an ordinary least square (**OLS**) linear model for statistical significance of these main factors and their 2-way interactions.

Part I: Predicting Hatchability

Feature Engineering Factors in the dataset potentially affecting hatchability were singled out and preprocessed. Month was a cyclical feature, from 1 to 12, meaning that the difference between the successive values is 1. However, there is always a jump from 12 to 1 to complete the cycle. By converting the months to their corresponding cosine and sine values, different coordinates were assigned for every moment between 1 and 12, making them unique. Furthermore, on the categorical factors (setters, hatchers, breeder farms, generation [grant-parent or parent stock], genetic strain) one-hot-encoding was applied to transform them into numeric values (0s and 1s). This process converted factors into forms that could be provided to machine learning (**ML**) algorithms for analysis. One-hot-encoding leads to an increased number of features, especially if there are many categories and this can make algorithms not to perform better. To solve this problem, dimensionality reduction was applied.

Dimensionality Reduction Dimensionality reduction was done by extracting new features from the original features, a process referred to as feature extraction

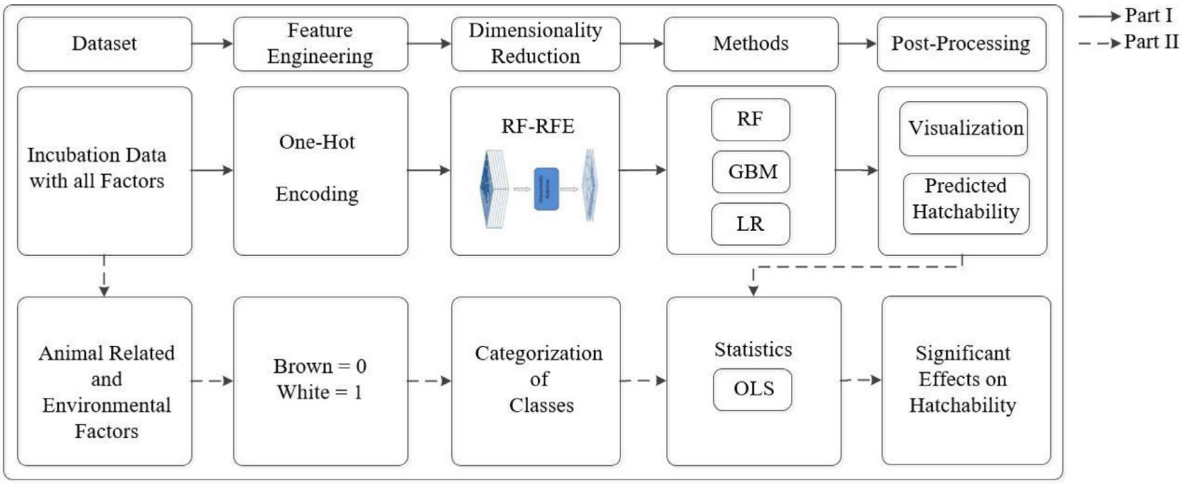


Figure 1. Flow chart, showing the methods used to predict hatchability (part I) and to investigate effects of animal related and environmental factors on hatchability (part II). Random forest (RF), gradient boosting machine (GBM) and linear regression (LR) methods were used to predict hatchability. Random forest - Recursive feature elimination (RF-RFE) was applied to reduce the dimensionality of the processed data to increase accuracy in predicting hatchability. The OLS (ordinary least square) was used to analyze animal related and environmental effects on hatchability.

(Osman et al., 2018). To perform feature extraction, a Recursive Feature Elimination with Cross Validation (**RFECV**) algorithm was applied on a Random Forest Classifier with a StratifiedKFold, K = 10 (Chen et al., 2018). The RFECV technique is recursive in the sense that for each feature, it calculates its importance score relative to the other features with hatchability as dependent variable. The algorithm helps to detect interactions between different features and classify them, but highly correlated features could mask these interactions. Thus, before fitting the model, Pearson's correlations (r) were calculated to eliminate highly correlated features with r >0.8. Consequently, only one feature from the generation (the PS) was eliminated, but the GPS remained.

Methods and Performance Evaluation To predict hatchability, the LR method and 2 independent ensemble ML methods, RF and GBM, were used. Analyses were performed in Python version 3.7.4 and the models were trained in the sklearn machine learning library. The methods were trained by splitting the dataset, using 60% as training set and 40% as test set to predict hatchability (Liu and Cocea, 2017).

For the 2 ensemble ML methods, RF and GBM, prior to training, hyperparameter optimization with K-Fold cross validation (**CV**) was performed, using a GridSearchCV algorithm. Hyperparameter optimization or tuning is a technique that is used to choose a set of optimal hyperparameters for a learning algorithm (Heaton, 2017). A hyperparameter is a parameter whose value is used to control the learning process of a ML algorithm (Ghawi and Pfeffer, 2019). Three options of hyperparameter values for the learning rate (0.001, 0.01, 1), number of estimators (500, 1,000, 10,000), just to name a few, were passed to the algorithms. This allowed each of the machine learning methods to autonomously choose from each option, which values were best to optimize predictive performance. Furthermore, to validate the stability of these models, K-Fold CV (K = 10) was applied to ensure the models got most of the patterns in the dataset (generalize). This automatically divided the dataset into 10 subsets, the hold out method was repeated 10-times such that 1 of the 10 subsets was used as the test/validation set and the other 9 (K-1) subsets were put together to form the training set. The error estimation was averaged over all 10 trials to get the overall effectiveness of the models. For the LR method, direct training was done without prior hyperparameter optimization, because LR does not support hyperparameter optimization. Finally, the predictive performance of the RF, GBM, and LR models were evaluated based on two standard performance metrices R^2 and RMSE.

Part II: Effects of Genetics and Environmental Factors on Hatchability

Effects of animal related (breeder strain, breeder age, egg weight uniformity) and environmental (egg storage duration, egg weight loss, season) factors on actual hatchability were evaluated. Breeder age was the age of the flocks in weeks and was categorized into 4 classes before analyses: <30 wk, 30 to 45 wk, 45 to 60 wk and >60 wk. Egg weight uniformity referred to homogeneity in egg weight at set within a batch (based on 150 eggs per batch) and was categorized into 3 classes: <85.0, 85.0 to 90.0% and >90.0%. These categories express the percentage of eggs within the average egg weight of the batch ± 10%. Egg storage duration was the number of days the eggs were stored at the breeder farm plus the hatchery prior to placement in the setters (Grochowska et al., 2019) and was categorized into 5 classes: 0 to 4 d, 5 to 7 d, 8 to 10 d, 11 to 14 d and 15 to 18 d. A storage temperature of 16 to 18°C was set when the eggs were stored for less than 7 d and when the storage period was longer, a temperature of 10 to 12°C was employed. Season was the period the eggs were set and was categorized into 4 classes (December, January, February = Winter; March, April, May = Spring; June, July, August = Summer; September, October, November = Autumn). Egg weight loss during incubation (from setting to transfer to the hatcher baskets) was categorized into 3 classes: <10.0%, 10.0 to 11.9% and ≥12.0%. Egg weight loss was calculated as the percentage of egg weight at transfer (from setter to hatcher on either day 17, 18, or 19 of incubation) relative to the egg weight prior to setting the eggs in the setter at the start of incubation. Breeder strain was categorized into brown and white, where both the brown and white strain included 11 genetic strains each. Data was analyzed, using a general linear model (Equation 2):

$$y = \mu + \mathrm{BA} + \mathrm{EWU} + \mathrm{ESD} + \mathrm{season} + \mathrm{EWL} + \mathrm{Strain} + \mathrm{interactions} + \mathrm{e}, \quad (2)$$

where y = Hatchability, μ = overall mean, **BA** = breeder age class (four levels), **EWU** = egg weight uniformity class (3 levels), **ESD** = egg storage duration class (five levels), season = season (spring, summer, autumn, winter), **EWL** = egg weight loss class (3 levels), strain = strain (white vs. brown), interactions = 2-way interactions between all these factors and e = residual error. Firstly, Pearson's correlations between these factors were calculated to ensure that they were not highly correlated. Preliminary analyses indicated that the strongest correlation between these features was r = 0.14 and none of them were significant. Consequently, these factors and their 2-way interactions were all included in the model. Thereafter, the model was reduced by stepwise deleting of the 2-way interactions. Main effects all remained in the model. A P-value of ≤ 0.05 was used as a threshold for significant main effects and interactions. Results are expressed as predicted means with SEM, and P-values obtained from the OLS statistical model. Bonferroni correction was used for multiple comparisons. All statistical analysis was performed in python version 3.7.4 using statsmodels, scipy, and scikit_posthocs libraries.

RESULTS AND DISCUSSION

Part I: Predicted Hatchability

Prior to predicting hatchability, dimensionality reduction was performed (Terko et al., 2019). Figure 2A shows that, out of the 125 features that were introduced to the algorithm, 103 were extracted as optimal features and were used to analyze hatchability. Figure 2B shows the top 8 features out of the 103 optimal features. Here, egg weight loss was the most important feature in predicting hatchability, meaning egg weight loss contributed the most in improving the performance (predictive accuracy) of the machine learning models.

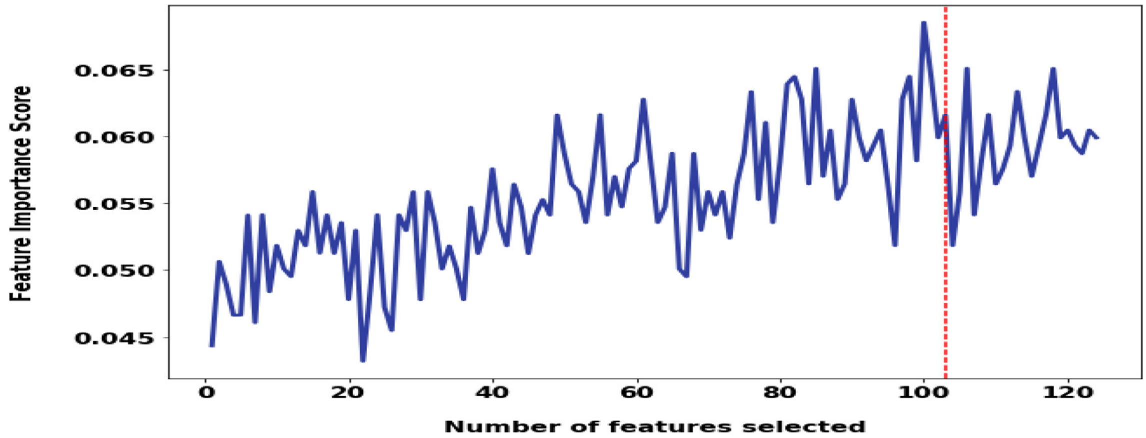
Figure 3 shows a scatter plot of predicted hatchability vs. the actual hatchability for RF, GBM, and LR models. The predictive performance based on RMSE for the RF, GBM and LR were 8.41, 8.67 and 8.92, respectively, whereas the predictive performance based on R^2 of the models were 0.35, 0.31, and 0.27, respectively. It appears that both ML (RF and GBM) models could fit the data in a similar and had a higher predictive accuracy than the LR model as indicated by the higher R^2 and lower RMSE, which is in agreement with studies on other biological processes (Wekesa et al., 2020). To further evaluate, again RF and GBM models showed consistently lower SEM than the LR model (see next section). The success of RF and GBM may be due to their ability to learn hidden interactions of features. The regression fits in Figure 3 showed deviation from the 45° angle for all 3 used models. This appears to be particularly related to the low values of actual hatchability, where higher values were predicted. This might be due to the relatively low number of observations in the dataset below hatchability of 60%. Increasing the number of observations would probably be the best solution to improve the performance of the models. The highest performance in terms of R^2 was recorded by the RF (0.35), which is still low to rely on. However, it also shows that ML models might perform better to predict biological processes with a lot of variation compared to LR models.

Part II: Genetic and Environmental Effects on Hatchability

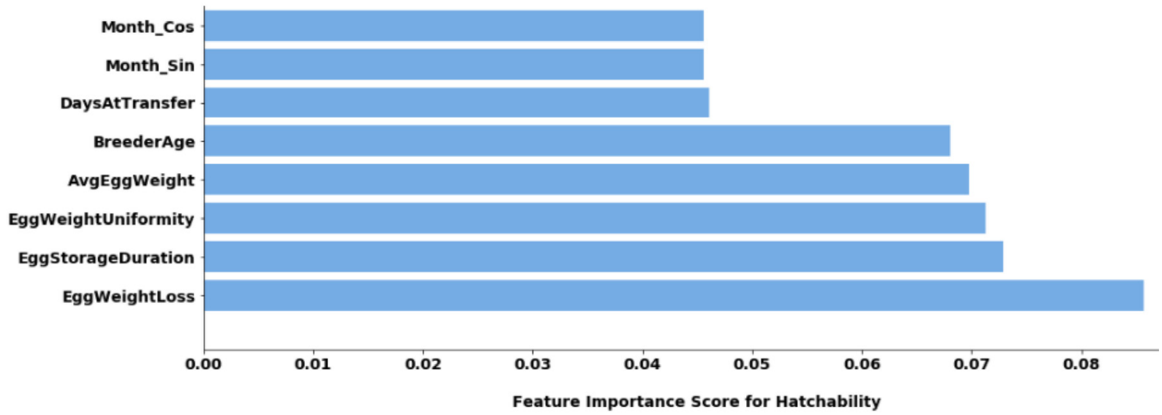
Five main effects and four 2-way interactions showed significant effects on actual hatchability. First animal related factors (strain, breeder age, egg weight uniformity) will be discussed, followed by environmental factors (egg storage duration, season). In this section, the figures express the actual hatchability, while the tables express the predicted hatchability by each model (LR, RF, GBM) and their overall P-values and SEM.

Effect of Strain on Hatchability Genetic strain had a highly significant effect on hatchability (P < 0.001). Strain also showed a significant 2-way interaction with breeder age and with season. Overall, the predicted hatchability of the brown genetic strain was higher than that of the white genetic strain (on average over the 3 models Δ = 2.02%; Table 1), which was similar with to (Narushin et al., 2016; Iqbal et al., 2016). This might be due to the differences in embryo physiology of the brown and white strains and higher egg weight in brown strains (Tona et al., 2010). Another reason for the differences in hatchability between the brown and white genetic strain might be due to the initial egg composition especially the distribution of yolk and albumen, which is determined by strain besides breeder age and egg weight (Van Der Wagt et al., 2020). Whether or not this difference in hatchability between strains is related to specific genes or single nucleotide polymorphism (**SNP's**) needs to be investigated further (Samiullah et al., 2015).

Effect of Breeder Age on Hatchability Breeder age and its interaction with strain had a highly significant effect on hatchability (P < 0.001). Figure 4 shows the relationship between breeder age and actual hatchability for white and brown strains. On average, white and



(a)



(b)

Figure 2. (A) Relationship between the number of features selected and the cumulative feature importance score, (B) top 8 of 103 selected features related to hatchability.

brown strains showed the highest hatchability at a breeder age between 30 and 45 wk of age. However, in the brown strain, hatchability remained rather constant and just declined after particularly 60 wk of age, whereas in the white strain the decline in the hatchability was more linear across the whole range of breeder ages. This might be related to differences in eggshell conductance between the 2 strains. It can be speculated that in older white leghorns the eggshell quality is decreasing faster than in brown strains. Consequently, more hair cracks or higher egg weight loss during incubation can occur, which might have a negative effect on hatchability. Thus, in white strains the hatchability will decrease, whereas this will less frequently occur in brown strains with a better eggshell quality. Table 2 shows the average hatchability per breeder age class for the 3 different models. The breeder age class of <30 wk of age had too limited observations to calculate predicted hatchability trustfully and consequently, this breeder age class was not taken into account. All 3 models showed the same trend, with the lowest hatchability after 60 wk of age. Eggs laid by old breeders often presented higher infertility and total embryo mortality, resulting in lower hatching percentage (Almeida et al., 2008). However, the difference in hatchability between a breeder age of 30 to 45 wk (highest hatchability) and above 60 wk (lowest hatchability) was considerably different between the 3 models (Δ = 5.54, 2.93 and 3.68% for LR, RF, and GBM, respectively). Comparing the variation between the highest and the lowest hatchability for the 3 models, the LR model showed the highest variation. This might be because the LR predicted hatchability values, which deviated more from the actual hatchability due to its inability to support parameter tuning prior to training, unlike the RF and GBM models.

Effect of Egg Weight Uniformity on Hatchability

Egg weight uniformity and its interaction with breeder age ($P < 0.001$ and $P = 0.04$, respectively) showed a significant effect on hatchability. Figure 5 shows the relationship between egg weight uniformity and actual hatchability for 3 breeder age classes. The breeder age class <30 wk did not have enough observations in the

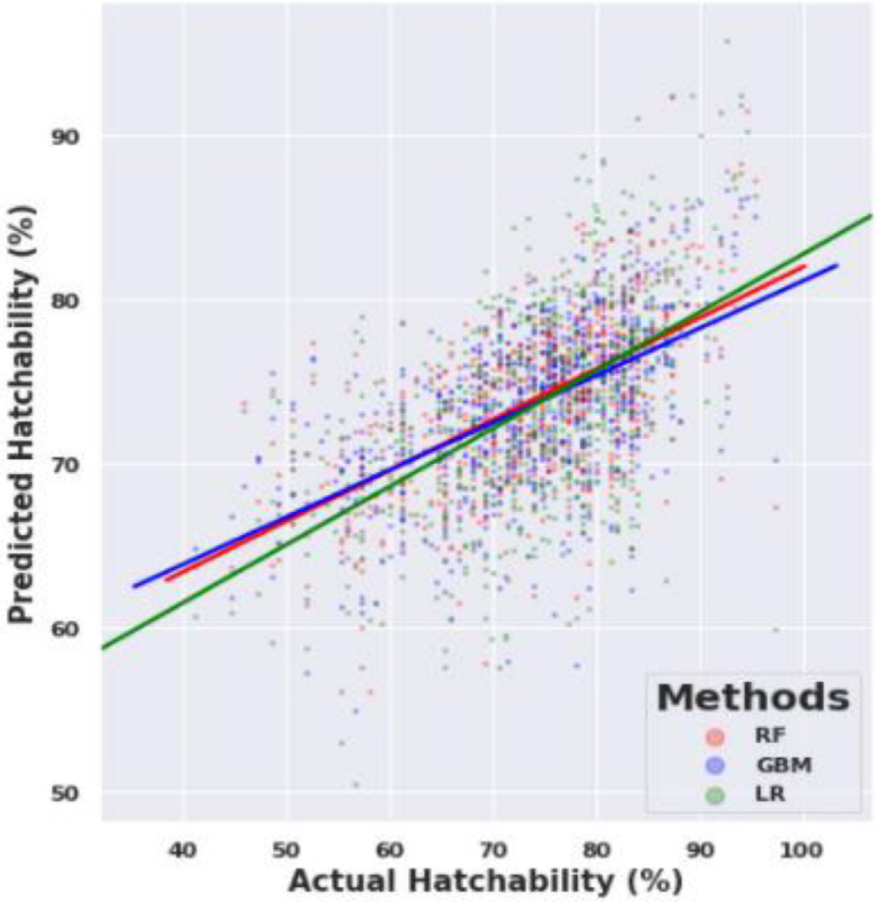


Figure 3. Regression plot of actual vs predicted hatchability of layer breeders for three models.

Table 1. Predicted hatchability for different levels of strain (grandparent and parent) by 3 models.

		Predicted hatchability (%)		
Strain	N*	Linear regression	Random forest	Gradient boosting machine
Brown	905	82.10[a]	82.29[a]	82.10[a]
White	789	79.86[b]	80.39[b]	80.15[b]
SEM		1.97	2.06	1.48
P-value		<0.001	<0.001	<0.001

*Number of batches.
[a,b]Predicted means within a column lacking a common superscript differ significantly ($P < 0.05$).

dataset to estimate a good relationship between egg weight uniformity and hatchability, thus it was not considered in the analyses. For breeders older than 45 wk, hardly any effect of egg weight uniformity on hatchability was found, but for breeders between 30 and 45 wk of age, higher egg uniformity resulted in a higher hatchability. In literature, more focus is given to broiler breeders and their body weight uniformity (Abbas et al., 2010) in relationship to hatchability. However, hardly any relationships between egg weight uniformity or the interaction between egg weight uniformity and breeder age on hatchability have been shown. It can be speculated that in younger breeders, variation in egg weight within a batch of eggs varies more than in older breeders, which consequently might have effect on heat transfer within an incubator. For example, Elibol and Brake (2008) demonstrated that eggshell temperature of large broiler breeder eggs during incubation was more affected by the place in the incubator (near the fan or far away from the fan) than that of smaller eggs. Consequently, it can be speculated that a batch of eggs with more variation in egg weight experience more variation in eggshell weight, affecting hatchability. Additionally, younger broiler breeders produce smaller eggs and it has been shown that hatchability of smaller eggs is lower compared to that of medium and large eggs (Iqbal et al., 2016). Table 3 shows the average hatchability per egg weight uniformity class for the 3 different models. All 3 models showed a similar trend, with the lower hatchability recorded for egg weight uniformity <85% compared to the other egg weight uniformity classes. The difference in hatchability between an egg weight uniformity >90% (highest hatchability) and <85% (lowest hatchability) was not considerably different between the 2 ML models (RF and GBM), but differed between the ML models and the LR model (Δ = 3.54, 2.32 and 2.85% for LR, RF, and GBM, respectively). This might be because RF and GBM are both ensemble (use a combination of tree algorithms) methods to do prediction unlike the LR method.

Effect of Egg Storage Duration on Hatchability Egg storage duration and its interaction with breeder age showed a highly significant effect on hatchability ($P <$ 0.001). Figure 6 shows the relationship between egg

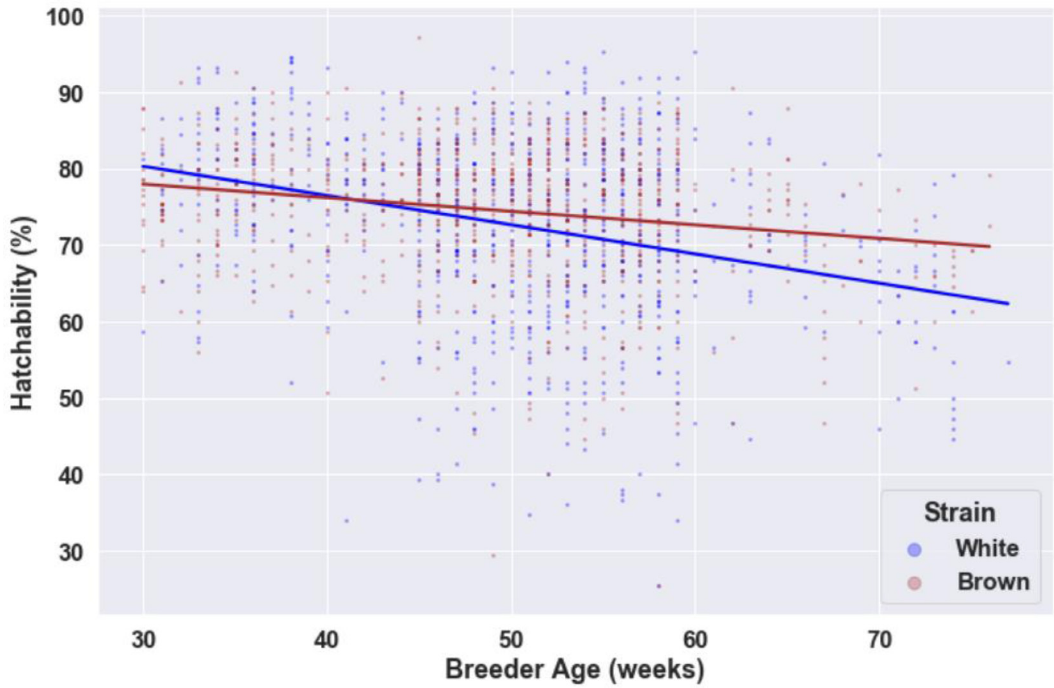


Figure 4. Relationships between breeder age (BA) and hatchability of white and brown layer breeder strains (grandparent and parent stock). Each dot represents a batch of eggs. $\textit{Hatchability} = 97.62 - 4.45 * \text{BA} + 0.98(\text{BA} * \text{Strain})$.

Table 2. Predicted hatchability for different levels of breeder age (grandparent and parent) by 3 models.

		Predicted hatchability (%)		
Breeder age (wk)	N*	Linear regression model	Random forest model	Gradient boosting machine model
30−45	398	82.53 [a]	82.74[a]	82.69[a]
45−60	1131	80.99[b]	81.03[b]	80.85[b]
>60	158	76.99[c]	79.81[c]	79.01[c]
SEM		0.53	0.56	0.40
P-value		<0.001	<0.001	<0.001

*Number of batches.
[a,b,c]Predicted means within a column lacking a common superscript differ significantly ($P < 0.05$).

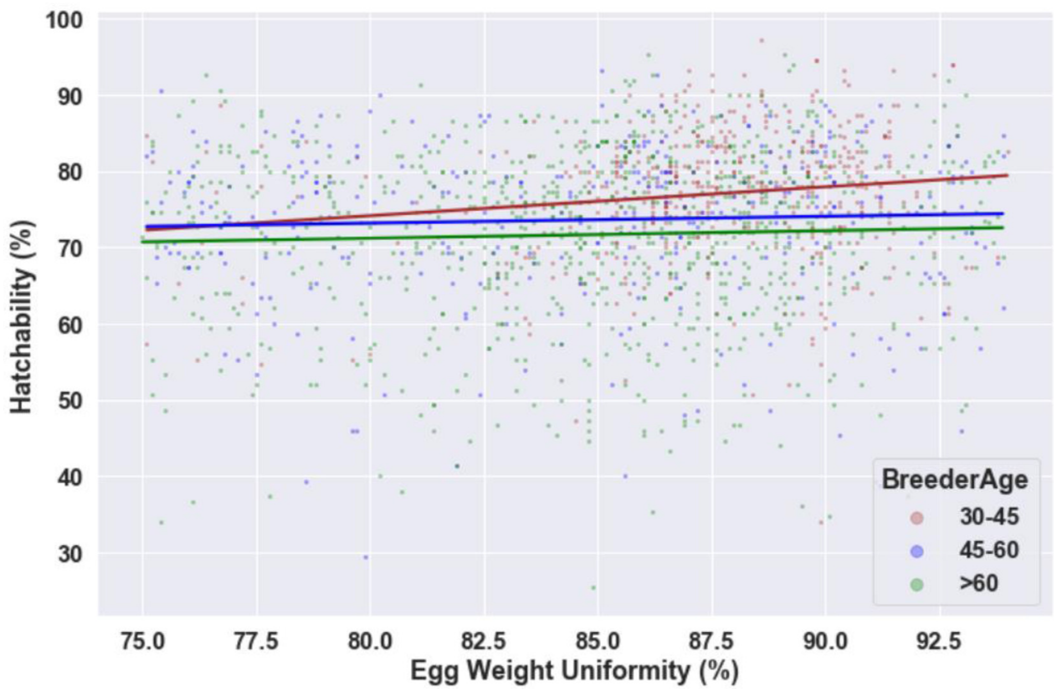


Figure 5. Relationships between egg weight uniformity (EWU) and hatchability of three layer breeder age (BA) classes, grandparent and parent stock. Each dot represents a batch of eggs. $\textit{Hatchability} = 97.62 - 0.33 * \text{EWU} + 0.34(\text{EWU} * \text{BA})$.

Table 3. Predicted hatchability for different levels of egg weight uniformity (grandparent and parent) by 3 models.

		Predicted hatchability (%)		
Egg weight uniformity(%)	N*	Linear regression model	Random forest model	Gradient boosting machine model
<85	584	79.45[c]	79.63[b]	79.43[b]
85−90	824	81.26[b]	82.27[a]	81.84[a]
>90	286	82.98[a]	81.95[a]	82.28[a]
SEM		0.31	0.33	0.24
P-value		<0.001	<0.001	<0.001

*Number of batches.
[a,b,c]Predicted means within a column lacking a common superscript differ significantly ($P < 0.05$).

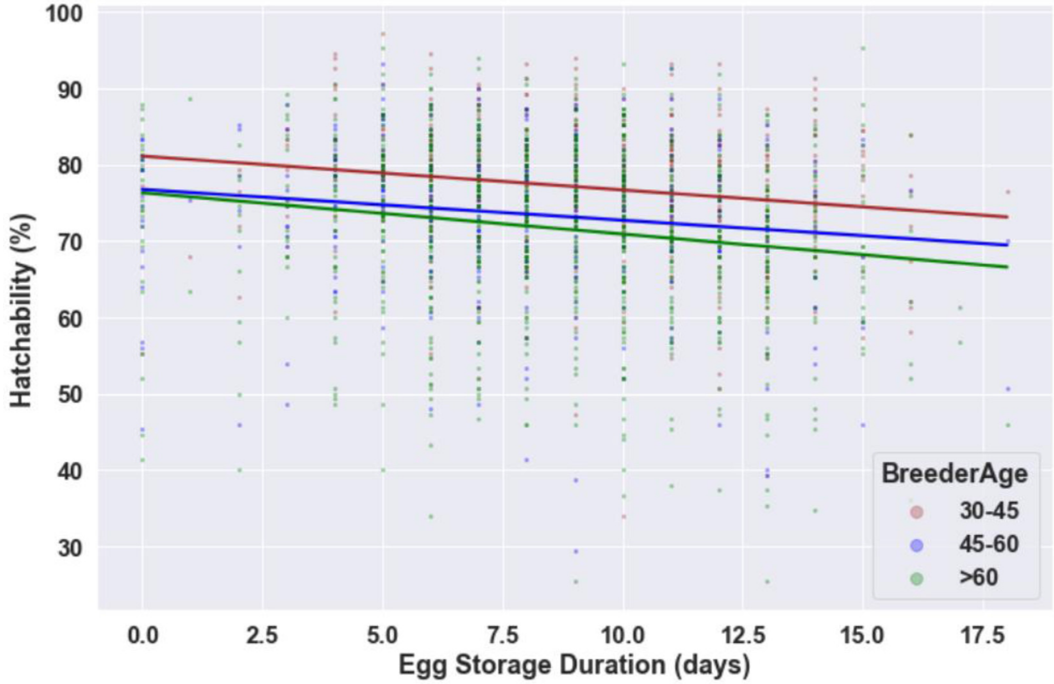


Figure 6. Relationships between egg storage duration (ESD) and hatchability of 3-layer breeder age (BA) classes: 30−45, 45−60, and >60 wk of age (grandparent, and parent stock). Each dot represents a batch of eggs. ***Hatchability*** $= 97.62 - 2.45 * \text{ESD} + 0.34(\text{ESD} * \text{BA})$.

Table 4. Predicted hatchability for different levels of egg storage duration (grandparent and parent) by 3 models.

		Predicted hatchability (%)		
Egg storage duration (d)	N*	Linear regression model	Random forest model	Gradient boosting machine model
0−4	185	83.24[a]	82.06[a]	82.02[ab]
5−7	474	82.12[a]	82.15[a]	82.18[a]
8−10	532	80.61[b]	81.64[a]	81.42[b]
11−14	446	79.45[c]	80.13[b]	79.61[c]
15−18	57	78.07[c]	76.88[c]	76.30[d]
SEM		0.23	0.24	0.17
P-value		<0.001	<0.001	<0.001

*Number of batches.
[a,b,c,d]Predicted means within a column lacking a common superscript differ significantly ($P < 0.05$).

storage duration and actual hatchability for the breeder age classes (30−45, 45−60 and >60 wk). All 3 breeder age classes showed the highest hatchability at storage duration of approximately 5 to 7 d (Silva et al., 2008). With prolonged storage duration, the hatchability declined for all breeder age classes, but this decline was more severe in the breeders > 45 wk than in the prime flock (30−45 wk) breeders. This also confirms what is shown in broiler literature: egg storage duration longer than 7 d, especially from older breeders, results in modifications to the blastoderm, which has a negative effect on hatchability (Damaziak et al., 2021). Adapted storage conditions related to the age of breeders might be an option to reduce negative effects of prolonged storage on hatching egg quality (Nasri et al., 2020c). Table 4 shows the average hatchability per egg storage duration class for the 3 different models. In general, all 3 models showed the same trend, with the highest hatchability for egg storage lower than 7 d. In particular the 2 ML models (RF and GBM) could specifically predict the highest hatchability at egg storage of 5 to 7 d due to their high ability to quickly learn a pattern in the dataset. The

Table 5. Predicted hatchability for different levels of egg weight loss (grandparent and parent) by 3 models.

Classes of egg weight loss (%)	N*	Predicted hatchability (%) Linear regression model	Random forest model	Gradient boosting machine model
<10	555	81.50[a]	81.49	81.34[a]
10.0−11.9	783	81.07[a]	81.58	81.39[a]
≥12.0	356	79.79[b]	80.36	80.02[b]
SEM		0.06	0.06	0.04
P-value		<0.001	0.27	0.02

*Number of batches.
a,bPredicted means within a column lacking a common superscript differ significantly ($P < 0.05$).

lowest hatchability was after 15 to 18 days of egg storage. However, the difference in hatchability between egg storage duration class 5 to 7 d (highest hatchability) and 15 to 18 d (lowest hatchability) was considerably different between the 3 models (Δ = 4.05, 5.27 and 5.88% for LR, RF, and GBM, respectively).

Effects of Egg Weight Loss on Hatchability Table 5 shows the average predicted hatchability per egg weight loss for the 3 different models. Based on the RF model, again there was no effect of egg weight loss on hatchability. However, LR and GBM models showed that egg weight loss had a significant effect ($P < 0.001$ and $P = 0.02$, respectively) on hatchability. This discrepancy between the models could be because of lower predictive performance of LR and GBM compared to the RF as indicated by their lower R^2 and higher RMSE. Previous studies have shown that there is little or no effect of egg weight loss on hatchability (Wolc et al., 2010; Hossain et al., 2017) and differences in egg weight caused by oviposition times does not have an effect on embryo development (Akil and Zakaria., 2015).

Effect of Season on Hatchability Season ($P = 0.03$) and its interaction with strain ($P < 0.001$) showed a significant effect on hatchability. Figure 7 shows the relationship between set month (the month the eggs were transferred into the setter trays) and actual hatchability for 2 strains (white vs. brown). For all the months, the hatchability of the brown strain was higher than that of the white strain. Table 6 shows the average hatchability per season for the 3 different models. Only the LR model showed an effect of season on hatchability, whereas both ML models did not. Based on the LR model, the hatchability was highest in Spring followed by Winter, Autumn and lastly by Summer. A comparable effect of season on hatchability was also shown by (Yassin et al., 2008) in broilers and (Jesuyon and Salako, 2013) in layers. The latter study demonstrated the influence of Early Wet (April−July), Late Wet (August−October) and Early Dry (November−January) and Late Dry (February−March) seasons (of humid climate) on hatchability of Bovan Nera (**BN**) and Isa Brown (**IB**)

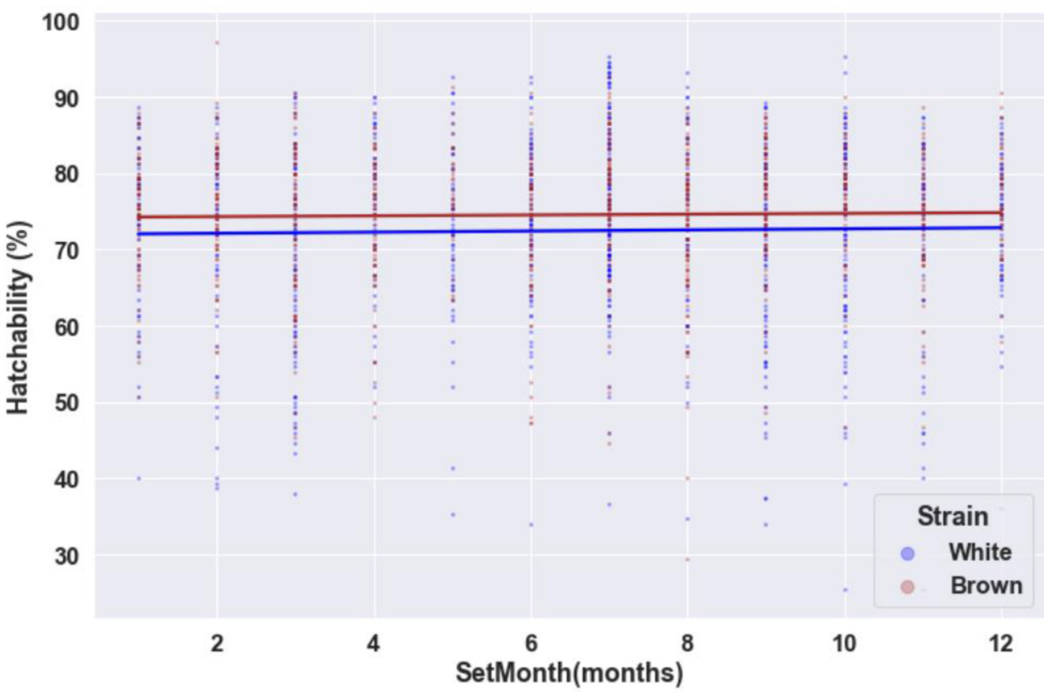


Figure 7. Relationships between month of set (the month the eggs were placed in the setter trays) and hatchability of two-layer breeder strains (white vs. brown) of grandparent and parent stock. Each dot represents a batch of eggs. $Hatchability = 97.62 + 0.28 * \mathrm{SetMonth} - 0.26(\mathrm{SetMonth} * \mathrm{Strain})$.

Table 6. Predicted hatchability for different levels of seasons (grandparent and parent) by 3 models.

		Predicted hatchability (%)		
Seasons	N*	Linear regression model	Random forest model	Gradient boosting machine model
Spring	360	81.70[a]	81.64	81.36
Summer	526	80.26[b]	81.38	81.05
Autumn	455	80.98[ab]	81.28	81.00
Winter	353	81.18[ab]	80.92	81.02
SEM		0.16	0.17	0.12
P-value		0.03	0.84	0.63

*Number of batches.
[a,b]Predicted means within a column lacking a common superscript differ significantly ($P < 0.05$).

parent-stock eggs. In Bovan Nera and Isa Brown, season had significant influence on egg fertility and egg hatchability, respectively (Jesuyon et Salako., 2013).

In can be concluded that, animal related and environmental factors are important factors in predicting hatchability of layer breeder eggs. It has been shown that machine learning models (RF, GBM) could predict hatchability more accurately than the LR model. All the 3 models showed consistency with each other in predicting and analyzing hatchability but predicted means may differ between the 3 models. Breeder age, egg storage duration, strain and egg weight uniformity had the strongest effects on hatchability, whereas season and egg weight loss only showed minor or no effects. Using large numbers field data, factor analyses more or less confirms what experimental data already found and the advantage of ML is that all factors are used in the models. This also allowed comparison of the best explaining factors for Hatchability.

ACKNOWLEDGMENTS

This study as part of the "Chicken Stress European Training Network" project financed by Marie Skłodowska-Curie is gratefully acknowledged. The authors would like to thank Hendrix Genetics for providing the data, especially the hatchery manager Bert Peters.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie [grant number 812777].

Disclaimer: This work reflects only the author's view and that the Agency European Union's Horizon 2020 research and innovation program is not responsible for any use that may be made of the information it contains.

DISCLOSURES

All authors declare that they have no conflict of interest.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.psj.2021.101394.

REFERENCES

Abbas, S. A., A. A. Gasm Elsei, and M.-K. A. Ahmed. 2010. Effect of body weight uniformity on the productivity of broiler breeder hens. Int. J. Poult. Sci. 9:225–230.

Almeida, J. G., S. L. Vieira, R. N. Reis, J. Berres, R. Barros, A. K. Ferreira, and F. V. F. Furtado. 2008. Hatching distribution and embryo mortality of eggs laid by broiler breeders of different ages. Braz. J. Poult. Sci. 10:89–96.

Akil, R., and A. H. Zakaria. 2015. Egg laying characteristics, egg weight, embryo development, hatching weight and post-hatch growth in relation to oviposition time of broiler breeders. Anim. Reprod. Sci. 156:103–110.

Chen, Q., Z. Meng, X. Liu, Q. Jin, and R. Su. 2018. Decision variants for the automatic determination of optimal feature subset in RF-RFE. Genes 9:309.

Coronel-Reyes, J., I. Ramirez-Morales, E. Fernandez-Blanco, D. Rivero, and A. Pazos. 2018. Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques. Comput. Electron. Agric. 145:1–10.

Damaziak, K., M. Koznaka-Lipka, D. Gozdowski, A. Gołębiowska, and E. Kędziorek. 2021. Effects of broiler breeder strain, age, and eggs preheating profile in single-stage systems on the hatchability of eggs and quality of chicks. Animal 15:100057.

Duman, M., and A. Şekeroğlu. 2017. Effect of egg weights on hatching results, broiler performance and some stress parameters. Braz. J. Poult. Sci. 19:255–262.

Dymond, J., B. Vinyard, A. D. Nicholson, N. A. French, and M. R. Bakst. 2013. Short periods of incubation during egg storage increase hatchability and chick quality in long-stored broiler eggs. Poult. Sci. 92:2977–2987.

Elibol, O., and J. Brake. 2008. Effect of egg weight and position relative to incubator fan on broiler hatchability and chick quality. Poult. Sci. 87:1913–1918.

Ghawi, R., and J. Pfeffer. 2019. Efficient hyperparameter tuning with grid search for text categorization using kNN approach with BM25 similarity. Open Comput. Sci. 9:160–180.

Grochowska, E., A. Kinal, Z. Sobek, I. Siatkowski, and M. Bednarczyk. 2019. Field study on the factors affecting egg weight loss, early embryonic mortality, hatchability, and chick mortality with the use of classification tree technique. Poult. Sci. 98:3626–3636.

Heaton, J. 2017. An empirical analysis of feature engineering for predictive modeling. arXiv 2:1701.

Hossain, T., A. Hossain, A. Roy, M. Azad, and M. Hawlider. 2017. Effect of moisture loss on the hatchability of chicken, duck and quail eggs. J. Environ. Sci. Nat. Resour. 9:105–108.

Iqbal, J., S. H. Khan, N. Mukhtar, T. Ahmed, and R. A. Pasha. 2016. Effects of egg size (weight) and age on hatching performance and chick quality of broiler breeder. J. Appl. Anim. Res. 44:54–64.

Jesuyon, O. M. A., and A. E. Salako. 2013. Effect of seasons on the reproductive performance of Bovan Nera and Isa brown parent-stock chickens in a hot humid environment. Int. J. Anim. Vet. Adv. 5:212–215.

King`ori, A. M. 2011. Review of the factors that influence egg fertility and hatchability in poultry. Int. J. Poult. Sci. 10:483–492.

Klein, D. R., M. M. do Vale, M. F. R. da Silva, M. F. Kuhn, T. Branco, and M. P. dos Santos. 2020. Data mining as a hatchery process evaluation tool. Sci. Agric. 77:4.

Liu, H., and M. Cocea. 2017. Semi-random partitioning of data into training and test sets in granular computing context. Granular Comput. 2:357–386.

Machado, J. P., M. A. Mesquita, M. B. Café, S. D. Assis, S. Veríssimo, R. R. Santos, N. S. M. Leandro, and I. C. S. Araújo. 2020. Effects of breeder age on embryonic development, hatching results, chick quality, and growing performance of the slow-growing genotype. Poult. Sci. 99:6697–6704.

Narushin, V. G., V. P. Bogatyr, and M. N. Romanov. 2016. Relationship between hatchability and non-destructive physical measurements of chicken eggs. J. Agric. Sci. 154:359–365.

Nasri, H., H. van den Brand, T. Najar, and M. Bouzouaia. 2020a. Interactions between egg storage duration and breeder age on selected egg quality, hatching results, and chicken quality. Animals 10:1719.

Nasri, H., H. van den Brand, T. Najjar, and M. Bouzouaia. 2020b. Egg storage and breeder age impact on egg quality and embryo development. J. Anim. Physiol. Anim. Nutr. 104:257–268.

Nasri, H., H. van den Brand, T. Najjar, and M. Bouzouaia. 2020c. Interactions between egg storage duration and broiler breeder age on egg fat content, chicken organ weights, and growth performance. Poult. Sci. 99:4607–4615.

Osman, H., M. Ghafari, and O. Nierstrasz. 2018. The impact of feature selection on predicting the number of bugs. arXiv 1:1807.

Pokhrel, N., E. Ben-Tal Cohen, O. Genin, M. Ruzal, D. Sela-Donenfeld, and Y. Cinnamon. 2018. Effects of storage conditions on hatchability, embryonic survival and cytoarchitectural properties in broiler from young and old flocks. Poult. Sci. 97:1426–1440.

Ren, G., T. Lin, Y. Ying, G. Chowdhary, and K. C. Ting. 2020. Agricultural robotics research applicable to poultry production. A review. Comput. Electron. Agric. 169:105216.

Samiullah, S., J. R. Roberts, and K. Chousalkar. 2015. Eggshell color in brown-egg laying hens- a review. Poult. Sci. 94:2566–2575.

Silva, F., D. Faria, K. Torres, D. F. Filho, A. Coelho, and V. Savino. 2008. Influence of egg pre-storage heating period and storage length on incubation results. Braz. J. Poult. Sci. 10:17–22.

Tona, K., K. Agbo, B. Kamers, N. Everaert, H. Willemsen, E. Decuypere, and M. Gbeassor. 2010. Comparison of Lohmann white and Lohmann brown strains in embryo physiology. Int. J. Poult. Sci. 9:907–910.

Terko, A., E. Žunić, D. Đonko, and A. Dželihodžić. 2019. Credit scoring model implementation in a microfinance context. ICAT 27:1–6.

Van Der Wagt, I., I. C. De Jong, M. A. Mitchell, R. Molenaar, and H. Van Den Brand. 2020. A review on yolk sac utilization in poultry. Poult. Sci. 99:2162–2175.

Wekesa, J. S., J. Meng, and Y. Luan. 2020. A deep learning model for plant lncRNA-protein interaction prediction with graph attention. Mol. Genet. Genom. 295:1091–1102.

Wolc, A., I. M. S. White, W. G. Hill, and V. E. Olori. 2010. Inheritance of hatchability in broiler chickens and its relationship to egg quality traits. Poult. Sci. 89:2334–2340.

Yassin, H., A. G. J. Velthuis, M. Boerjan, J. van Riel, and R. B. M. Huirne. 2008. Field study on broiler eggs hatchability. Poult. Sci. 87:2408–2417.