

Reducing the environmental footprint in hatcheries through a new approach to sexing bird eggs

Alexander Aleynikov^{1,2*}, Irina Yakovina², and Ivan Osipenko²

¹Siberian Federal Scientific Centre of Agro technology, Russian Academy of Science, Novosibirsk region, Krasnoobsk, 630501, Russia

²Novosibirsk State Technical University, 20, Karl Marx Avenue, Novosibirsk, 630073, Russia

Abstract. Industrial poultry farming can satisfy the population's need for meat up to 98%, and for eggs – 92%. With the growth of world production of poultry products, the volume of hatchery waste also increases, because the hatched cockerel chicks are destroyed after incubation due to the inefficiency of their further cultivation (more than 7 billion). Determination of the sex of the embryo in the egg before incubation will significantly reduce the cost of egg production and the environmental burden from the activities of poultry farms. Within the framework of this article, the tasks of developing models for determining the sex of an embryo in a bird egg before incubation using machine learning (ML) methods are solved. During the first experiment, the identifiability of each of the samples was checked by the ML methods. During the second experiment, using various methods (decision trees, random forests, adaptive boosting, logistic regression and support vectors), a preliminary set of models was obtained. The third experiment ended with the formation of the resulting set of features and obtaining the final ML model. This made it possible to determine the sex of the embryo using 16 geometric parameters of the egg with an acceptable level of accuracy.

1 Introduction

Poultry farming is one of the developed traditional livestock industries in the world, because it provides food security for the inhabitants of many countries [1].

For example, the industrial poultry farming of domestic chickens can satisfy the population's need for meat up to 98%, and for eggs – 92% [2]. Indeed, the global production of poultry meat is about 140 million tons [3]. The intensification of the production of broilers, to obtain meat, and laying hens, to obtain a dietary product – eggs, has a significant impact not only on human health, but also on the environment [4].

As global poultry production grows, so does the amount of hatchery waste. There is a growing amount of egg shells and fluff, infertile eggs, dead embryos, culled chickens,

* Corresponding author: fti2009@yandex.ru

embryonic fluid, as well as wastewater obtained during the cleaning and disinfection of equipment and growing areas [3].

Due to the gender orientation of poultry production, every year in the world more than 7.0 billion one-day-old males are subject to destruction by barbaric methods – by maceration and suffocation in a carbon dioxide environment [6].

In 2018, the company «In Ovo» introduced the commercial robotic invasive technology Ella for sex determination in eggs in many European countries [7].

However, this technology is invasive, complex and expensive to implement.

Numerous attempts to determine sexual dimorphism based on egg shape coefficient or index have also not been successful [8].

Determination of the sex of the embryo in the egg before incubation is an unresolved world problem [7, 8]. The solution to this problem will not only remove ethical problems in society, but will significantly reduce the costs of egg production and environmental stress from the activities of poultry farms.

In our work, we develop a hypothesis about the different nature of the asymmetry of egg shape parameters in male and female embryos in a freshly laid egg of poultry [7], as well as the possibility of determining sexual dimorphism in it before incubation using modern ML.

2 Materials, data and methods

2.1 Data description

The experimental batch of samples consisted of 80 eggs of the Hisex White cross. Out of these, 38 chicks successfully hatched from the incubation process. The genders of the chicks were determined through visual inspection, resulting in the identification of 24 roosters and 14 hens.

Thus, the original samples were based on the characteristics of 38 egg images, for which the gender of the chick was identified with a certain level of confidence. These images were utilized to form a set of numerical geometric characteristics that were obtained through the processing and analysis of the source images using computer vision techniques and the specifically designed program. The original dataset consisted of 38 samples; each described by 93 features obtained using various image processing methods.

To build models empirically, a statistical analysis was conducted to test hypotheses about the differences in means for the studied groups.

However, the use of statistical analysis did not produce the expected results.

Therefore, a decision was made to investigate the applicability of machine learning methods for obtaining models that can determine the gender of a chick based on the geometric characteristics of the egg and identification of the most informative features.

When building models, the set of features obtained from various image processing methods were grouped into the several distinct categories:

- G01 (6 features) Basic characteristics: mass, perimeter, area, longitudinal and transverse dimensions, and overall shape index;
- G02 (11 features) Shape index based on segmental transverse and longitudinal dimensions;

Characteristics of the radius-vectors drawn from the center of the object to the contour boundary:

- G03 (36 features) Vector length by segment;
- G04 (18 features) Length ratio at an angle of 180 degrees;
- G05 (18 features) Length ratio at an angle of 90 degrees;

Characteristics of the completed image:

- G06 (16 features) – Group 1: area, perimeter;
- G07 (8 features) – Group 2: with the use of an inscribed circle;
- G08 (12 features) – Group 3: averaged characteristics;
- G09 (4 features) Calculated volume;
- G10 (93 features) Combination of groups G01-G09;
- G11 (53 features) Combination of groups G01-G03.

2.2 Methods, ML algorithms and tools used

Different machine learning algorithms were used to build the models such as decision trees, random forests, adaptive boosting, logistic regression, and support vector machines (SVM) for classification using the following configurations:

- M01: Decision Tree Classifier with max depth 3;
- M02: Decision Tree Classifier with max depth 5;
- M03: Random Forest Classifier with 4 estimators and max depth 3;
- M04: Random Forest Classifier with 10 estimators and max depth 5;
- M05: AdaBoost Classifier with 4 Decision Tree estimators and max depth 3;
- M06: Support Vector Classifier with "RBF" kernel;
- M07: Support Vector Classifier with "Linear" kernel;
- M08: Logistic regression with L2 regularization.

The Python programming language was utilized for data processing and analysis. The scikit-learn library was selected to implement the listed algorithms for machine learning models creation and training.

The Orange 3 program with a graphical user interface was used to build the models.

For evaluating the model's metrics, a cross-validation approach was employed in two different variations [9].

At the initial stage of the work, a leave-one-out control method was utilized, where the testing subset consisted of a single sample ($k=1$), and the number of divisions and models were equivalent to the number of samples ($N=38$).

In the model building and selection phase, a K-fold control method was implemented with $K=3$ partitions and averaging of the results.

The advantage of using cross-validation is that it allows for a more robust evaluation of the model's performance, as it tests the model's ability to generalize to new data that was not seen during training. This helps prevent overfitting, where the model performs well on the training data but poorly on new data.

Thus, cross-validation can provide a more accurate estimate of a model's performance and make it more reliable for practical use.

Among the numerous metrics available to evaluate classification models [10], this study utilized the AUC ROC (area under the receiver operating characteristic curve) and F1-measure. F1-measure is computed as the harmonic mean of precision and recall, providing equal weighting to both precision and recall.

3 Experiments

The versatility of the task at hand necessitated a series of experiments, each of which yielded the desired result.

During the first experiment, we tested the identifiability of each sample using machine learning methods.

The second experiment involved the use of various machine learning algorithms, which led to the creation of a preliminary set of models.

In the third and final experiment, we formed a resulting set of features and successfully obtained the final machine learning model.

Let us take a closer look at the progress of each of the conducted experiments.

3.1 Identifiability check

For datasets with small sample sizes, errors in data collection can have a significant impact on the final results.

In order to study the identifiability of objects from the original sample using machine learning models, the Leave-One-Out method was used with the application of ML algorithms M01-M06 for G01-G10 feature groups.

In the end, a total of $38 \times 10 \times 6$ models were built, and the experiment's results are presented in Figure 1 as a heat map.

Each cell of heat map shows the total number of correct conclusions out of 6 models obtained using different machine learning algorithms for an individual sample.

Feature group	Number of correctly recognized object and object gender																																						Sum	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38		
G01	6	4	2	2	3	4	3	6	2	2	6	4	5	2	6	6	6	5	5	5	6	5	5	2	2	1	0	2	1	0	0	1	3	0	0	1	0	2	115	
G02	6	3	2	6	6	6	6	6	2	4	5	4	6	3	4	6	3	5	2	6	4	6	5	4	0	4	0	2	1	0	3	3	1	0	0	0	1	2	127	
G03	6	2	2	2	6	6	6	5	1	3	2	6	4	3	5	5	5	5	5	6	2	6	5	2	0	2	3	0	1	0	2	1	1	0	1	2	2	2	117	
G04	6	4	6	5	4	6	6	6	6	4	4	6	6	2	2	4	3	5	6	5	2	3	5	5	1	1	0	0	1	1	0	1	2	2	0	1	0	2	123	
G05	5	6	2	2	6	5	3	6	2	6	6	6	6	2	6	6	6	6	2	6	3	6	3	5	1	4	0	1	0	0	3	2	3	3	0	2	0	2	133	
G06	6	5	2	3	6	6	6	4	2	2	3	3	5	6	5	6	6	4	4	4	5	5	6	2	2	2	0	1	1	1	0	0	0	1	0	1	0	1	116	
G07	6	2	1	3	6	5	5	6	2	5	3	6	6	1	6	6	6	6	5	6	6	6	1	3	0	1	2	0	1	1	0	1	2	5	1	0	0	1	123	
G08	6	1	1	6	5	6	6	6	2	4	2	6	6	1	6	6	6	5	5	6	6	6	6	2	3	0	0	2	1	0	0	3	1	0	4	3	0	0	1	124
G09	5	5	4	6	6	3	2	6	2	4	2	6	6	2	6	5	6	5	6	6	6	6	6	2	2	1	2	0	0	1	0	3	0	0	0	0	0	1	123	
G10	6	3	2	1	6	5	1	6	4	4	3	6	6	1	4	6	1	2	2	6	3	6	2	4	1	4	1	2	1	1	0	3	1	1	0	1	0	4	110	
	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	H	H	H	H	H	H	H	H	H	H	H

Fig. 1. Heat map displaying the result of the first experiment (R – Rooster, H – Hen).

Upon analyzing the heat map, it was found that some samples were identified by only a few generated models.

These samples exhibited low values for the sum of correct responses and were considered boundary values that excluded certain samples from the original dataset.

For this experiment, the following boundaries were selected: for rooster's ≤ 32 and for hen's ≤ 5 .

The following samples were removed from the original dataset due to difficulty in identification: roosters – 3, 9, 14, 24 and hens – 30, 35, 37.

3.2 Model formation

Following the exclusion of seven samples from the dataset, 31 samples remained – 11 hens and 20 roosters.

Methods M07 and M08 were added to the set of machine learning algorithms. Based on the results of the first experiment, groups G01, G02, G03, and G11 were selected for model formation.

For each feature group G01, G02, G03, G11, and machine learning algorithm M01-M08, models were built utilizing K-fold cross-validation with three partitions.

The metrics for each shuffle were averaged to obtain the final evaluation of classification metrics (Figure 2, 3). Overall, there were $4 \times 8 \times 3$ models generated.

The generalized results for AUC ROC and F1-measure are presented in Table 1. Among the obtained models, those with the highest averaged metrics (AUC=67-72%, F1=70-76%) were those generated by algorithms M04 and M05 utilizing features from groups G02 and G11. The metrics for feature group G02 (11 features) exceeded those from the models generated using the feature group G1 (53 features) so the models built with a smaller number of features produced better results than models with more features.

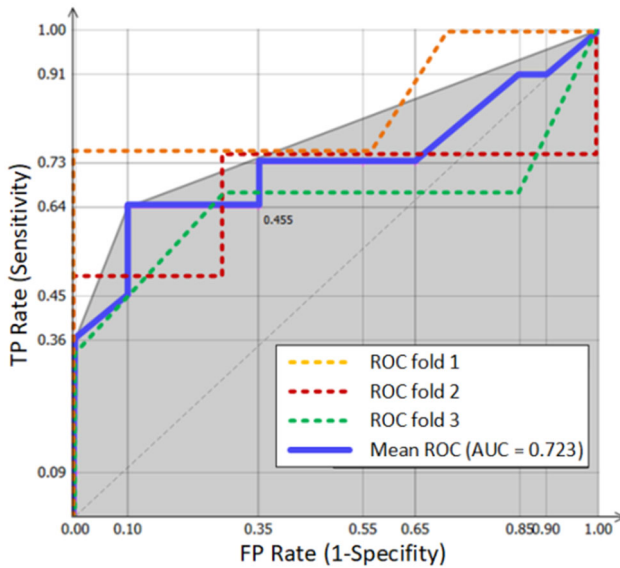


Fig. 2. Usage of ROC-analysis in first and third experiments: average ROC for 3-fold cross validation for single ML algorithm.

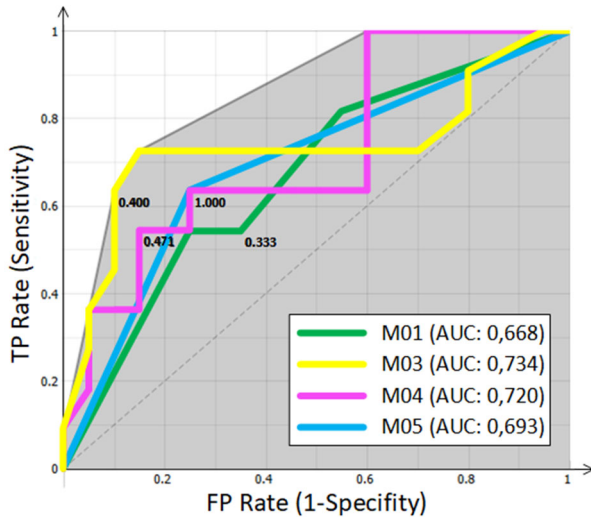


Fig. 3. Comparison of mean ROC curves: from model on G11 feature group.

Table 1. Result of experiments (Cross-validation, K=3).

№	Group of features	Avg. metrics	M01	M02	M03	M04	M05	M06	M07	M08
			2	G01	AUC	0.484	0.495	0.627	0.543	0.536
	F1	0.56	0.54		0.61	0.59	0.59	0.51	0.51	0.54
	G02	AUC	0.664	0.655	0.659	0.673	0.718	0.532	0.618	0.464
		F1	0.68	0.68	0.68	0.76	0.74	0.61	0.49	0.58
	G03	AUC	0.427	0.509	0.489	0.602	0.532	0.495	0.386	0.436
		F1	0.45	0.53	0.52	0.61	0.58	0.47	0.51	0.48
	G11	AUC	0.614	0.584	0.698	0.723	0.673	0.591	0.309	0.432
		F1	0.58	0.59	0.65	0.68	0.71	0.54	0.51	0.58
3	G12	AUC	0.668	0.634	0.734	0.720	0.693	0.600	0.591	0.364
		F1	0.68	0.62	0.73	0.70	0.71	0.51	0.58	0.49
	G13	AUC	0.602	0.627	0.668	0.702	0.602	0.577	0.386	0.368
		F1	0.63	0.65	0.67	0.61	0.64	0.51	0.51	0.54

It could be attributed to the curse of dimensionality, noise in the data, and an increase in entropy [11].

As the number of features increases, the amount of noise and randomness in the data also increases, making it more challenging to extract meaningful patterns and information from the data. Dimensionality reduction techniques and feature selection methods are useful tools to address these issues and improve model performance.

3.3 Model finalization

The goal of the third experiment was to enhance model metrics by using more informative features. This task was accomplished through implementing SHAP values [12].

The relative importance values were calculated for the best models of the second experiment – M04, M05, and feature group G11 (consisting of G01, G02, G03).

Using SHAP feature’s value reflecting, which reflect the relative importance of features two new feature groups was formed: G12 for model M04 and G13 for M05. Each group contains the 15 most informative egg geometric shape parameters (Figure 4).

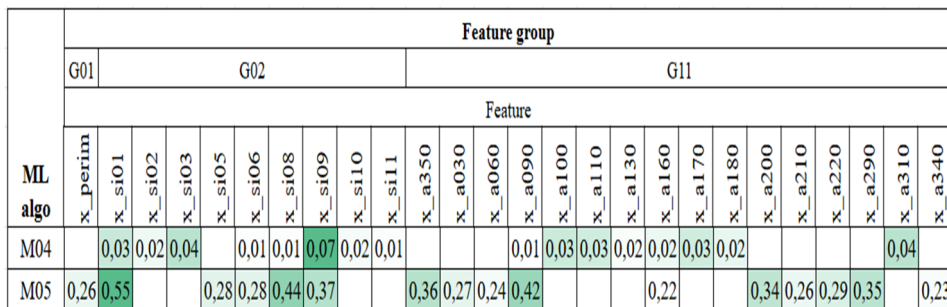


Fig. 4. Heat map of feature importance values obtained based on SHAP indices.

Using the newly truncated feature groups G12 and G13, all algorithms were retrained. As a result, the three final models – M03, M04, M05 using feature group G12, yielded the best results (see Table 1 and Figure 3 for third experiment results).

Thus, the obtained solution allows for determining avian embryo sex non-invasively before incubation with an acceptable level of metrics based on egg geometric shape parameters using machine learning methods.

4 Conclusions

As result of the proposed solution for the development of non-destructive models for determining the sex of embryos in bird eggs, three final models were obtained with accuracy metric values of AUC=73-72% and F1=69-72%: Random Forest Classifier with 4 estimators and max depth 3, Random Forest Classifier with 10 estimators and max depth 5, AdaBoost Classifier with 4 Decision Tree estimators and max depth 3.

The use of cross-validation to assess the accuracy metrics of the models helped reduce the impact of overfitting.

Reducing the feature set and selecting the significant 16 features allowed for an increase in AUC metrics up to 7% and F1 metrics up to 5% for some algorithms.

The resulting feature group included the egg's geometric characteristics: 8 out of 11 shape index parameters based on transverse and longitudinal dimensions by segments, and 8 out of 36 characteristics of radius vectors drawn from the center of the object to the contour boundary.

Thus, the obtained results indicate that the proposed combination of machine learning algorithms allows for the development of classification models for tasks where the feature space significantly exceeds the sample size with ambiguously identifiable samples.

These results make a certain contribution to solving the complex problem of sex determination in the egg embryo during the pre-incubation period.

Further research is planned to increase the amount of data and obtain classification models suitable for practical use.

The research was supported by the grant of the Russian Science Foundation No. 22-26-00198.

References

1. M. Vaarst, S. Steinfeldt, K. Horsted, *Worlds Poul. Sci. J.* **71**, 609-620 (2015)
2. A. Mottet, G. Tempio, *Worlds Poul. Sci. J.* **73**, 245-256 (2017)
3. G. Gržinić et al. *Sci. Total Environ* **858**, 160014 (2023)
4. M. De Vries, I. J. M. De Boer, *Livest. Sci.* **128**, 1-11 (2010)
5. E. F. Kaleta, T. Redmann, *Worlds Poultry Sci. J.* **64**, 391-399 (2008)
6. A. Khaliduzzaman, F. Shinichi, K. Ayuko, S. Tetsuhito, O. Yuichi, K. Naoshi, *Comp. Elect. Agric.* **158**, 30-35 (2019)
7. A. F. Aleynikov, *IOP Conf. Ser. Earth. Envir. Sci.* **1112**, 012057 (2022)
8. J. Rutkowska, A. Dubie, S. Nakagawa, *J. Evolut. Biol.* **27(1)**, 153-160 (2013)
9. T. T. Wong, *Pattern Recognition* **48(9)**, 2839-2846 (2015)
10. P. Branco, L. Torgo, R. P. Ribeiro, *ACM computing surveys (CSUR)* **49**, 1-50 (2016)
11. E. Debie, K. Shafi, *Aleyic.* **22**, 519-536 (2019)

12. W. E. Marcílio, M. E. Danil, *From explanations to feature selection: assessing SHAP values as feature selection mechanism*, in 33rd conf. on Graphics, Patterns and Images, SIBGRAPI, 07-10 November 2020, Porto de Galinhas, Brazil, 340-347 (2020)