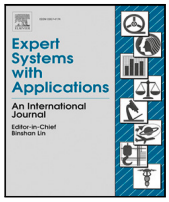
Expert Systems With Applications 238 (2024) 122130

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



The verification of hen egg types by the classification of ultra-weak photon emission data

Joanna Sekulska-Nalewajko [a,*], Jarosław Gocławski [a], Ewa Korzeniewska [b], Paweł Kiełbasa [c], Tomasz Dróżdż [c]

[a] Lodz University of Technology, Institute of Applied Computer Science, Stefanowskiego 18, 90-537 Łódź, Poland
[b] Lodz University of Technology, Institute of Electrical Engineering Systems, Stefanowskiego 18, 90-537 Łódź, Poland
[c] University of Agriculture in Cracow, Faculty of Production and Power Engineering, 116B Balicka Street, 30-149 Krakow, Poland

ARTICLE INFO

Keywords:
Egg quality
Photon emission
Support Vector Machine
Classification accuracy

ABSTRACT

Examples from living systems at various levels of the biological hierarchy and also from natural food products show that ultra-weak photon emission (UPE) has potential applications in the rating of vital functions and quality testing. In this study, the UPE of chicken eggs has been tested regarding the possibility of egg quality verification. The UPE from intact eggs and separated egg parts were subjected to supervised and unsupervised classification methods according to different housing types. The results of unsupervised egg grouping substantially agreed with the types of hen rearing. The Cohen's Kappa test score for the K-means method was up to $K = 0.63$. Supervised Support Vector Machine (SVM) classifier with radial kernel function achieved a relatively high accuracy (AC), up to 88%, also confirmed by the value of the K-statistics up to 0.81. This study shows that the best result of egg types classification can be obtained using UPE emission data from all egg parts.

1. Introduction

Chicken eggs are a staple food product, famous for their nutritional value and versatility. They are considered to be the perfect food due to their multi-functional properties. For economic and health reasons, it is crucial to inspect the quality of eggs.

There is much evidence in the literature that the yield and quality of eggs differ depending on the housing environments of the hens (Nain et al., 2012; Özbey & Esen, 2007). It was noted higher egg quality in domestic conditions than in conventional cages (Meng et al., 2014; Yenice et al., 2016) However, differences in the nutritional content of eggs produced using different farming techniques are not clearly delineated and challenging to identify.

The nutritional composition of eggs varies and depends on many factors, including the method of hen breeding, especially feed composition and access to herbaceous plants (the so-called green fodder) (McNamara, 2010; Nain et al., 2012; Özbey & Esen, 2007). Nowadays, it is relatively easy to modify the composition of the egg, especially the fatty acid content of the yolk, for example, by administering appropriate components to the laying hens in the feed. As a result, the level of certain ingredients can be significantly increased. It is often the case on organic farms, where laying hens have a much more varied ecological nutrition, with access to green fodder, many vegetables, herbs, and the possibility of natural supplementation with minerals in the diet (including digging in the ground, green forage, free-range) (McNamara, 2010; Yannakopoulos, 2007).

Traditional egg quality assessment methods, based on chemical analysis, are time-consuming and unsuitable for widespread use. Therefore, a faster method, which could be automated, would be an alternative to the possibility of assessing eggs already in the production process. Recent studies have shown that optical methods effectively evaluate egg quality and freshness (Brasil et al., 2022; Cruz-Tirado et al., 2021; Yao et al., 2022). These are mainly methods based on the spectral analysis of eggs, also operating in the near-infrared (NIR) range. Several authors have shown that by measuring the ultra-weak photon emission (UPE) from egg yolks, eggs can be differentiated regarding farming type (Grashorn & Egerer, 2007; Köhler, 2001; Köhler et al., 1991). Higher levels of photon emission may indicate higher-quality organic components resulting from better breeding conditions.

Ultra-weak photon emission is called ultra-weak luminescence (UWL) and delayed luminescence (DL). The term 'biophotons' is also commonly used, reflecting this phenomenon's nature. UPE is associated

* Corresponding author.
E-mail addresses: joanna.sekulska-nalewajko@p.lodz.pl (J. Sekulska-Nalewajko), jaroslaw.goclawski@p.lodz.pl (J. Gocławski), ewa.korzeniewska@p.lodz.pl (E. Korzeniewska), pawel.kielbasa@urk.edu.pl (P. Kiełbasa), tomasz.drozdz@office.urk.edu.pl (T. Dróżdż).

https://doi.org/10.1016/j.eswa.2023.122130
Received 30 January 2023; Received in revised form 2 October 2023; Accepted 10 October 2023
Available online 18 October 2023

primarily with the vital functions and biological activity of living organisms under normal conditions. All living systems constantly emit tiny amounts of photons per time unit, which can be detected only using sensitive photoelectron multipliers (Ruth & Popp, 1976). Ultra-weak luminescence is invisible to the human eye (in the range of visible light) because the number of photons emitted lies far below the absolute threshold for the energetic sensitivity of retinal cells, i.e. 10–103 photons per cm^2/s (Kobayashi, 2003). Photon emission is detected as non-thermal radiation, occurring in the electromagnetic spectrum in the range from near-ultraviolet to the visible region (100–800 nm) (Beloussov, 2003; Popp, 1992, 2003), and possibly reaching the near-infrared region (801–1300 nm) (Cifra & Pospišil, 2014). This radiation is also associated with dimol and monomol photon emission of $^1O_2$ in the infrared (≈1300 nm) area of the spectrum.

Ultra-weak luminescence can be expected in all biological systems in which biomacromolecules are available due to active redox/radical processes, passive oxidation by environmental oxidants, or the metabolic activity of microbial infestation in plant and animal products. This set of possible oxidative pathways makes UPE useful for examining food quality, including milk, rice, and bean sprouts (Iida et al., 2002). With food, UPE is used mainly to analyse the oxidation balance of organic components (Lambing, 1992).

Computer artificial classifiers have recently been studied intensively to support human decision-making for agricultural product quality (Nouri-Ahmadabadi et al., 2017). Eggs are also of interest in this type of classification. Mehdizadeh et al. (2014) used an artificial neural network (ANN) to build an intelligent system for chicken egg quality ranking based on intact egg visible–infrared transmittance. It is also an example of one of those studies that try to find a way to test egg quality non-invasive by spectroscopic techniques (Kemps et al., 2006; Liu et al., 2007).

In recent years, NIR spectroscopy data processed by machine learning methods have been successfully used for the prediction of eggs' quality (Brasil et al., 2022; Cruz-Tirado et al., 2021). In those studies, the classifications type Partial Least Squares Discriminant Analysis (PLS-DA), Support Vector Machine with Cost (SVM-C), K-means, or regressions type PLS-R or SVM-R were implemented to predict stale and fresh eggs. Sehirli and Arslan (2022) determined egg quality in haugh units (HU) based on 20 mechanical features of the egg parts (white, yolk, shell) and also a chicken label, genotype, colour, and family, based on a variety of prediction methods. Computer classifications were also applied to help egg fertility detection during incubation based on the detection of embryo presence. Such studies have been carried out for several years at the Universitas Pembangunan Nasional Veteran Yogykarta (Saifullah, 2020, 2021; Saifullah & Drezewski, 2022; Saifullah et al., 2022; Saifullah & Suryotomo, 2021). In these research both K-means segmentation and SVM classification were used for statistical features of egg images based on Gray Level Co-occurrence (GLCM) matrix — similar to those of Haralick et al. (1973).

This work presents the UPE application in an eggs type verification using unsupervised and supervised classifiers. The considered types of eggs relate to the method of laying hens breeding, and they are cage, free-range and organic breeding. Understanding the mechanism of ultra-weak photon emission from eggs and egg parts and quantifying its intensity, spatial and temporal distribution, and dependencies on internal biological activity and/or environmental conditions will help establish proper assumptions for UPE measurements and enable its application to egg quality assessment.

2. Materials and methods

2.1. Egg samples

All tested eggs, regardless of the type of hen rearing (cage, free-range, organic), were derived from the Rhode Island Red hen breed (R-11). The tested eggs came from farms located in one region of South Poland (Małopolska/near the city of Krakow) and were obtained at the turn of August and September. The eggs came from three farms with different rearing systems in line with the Commission Regulation. One-third of the eggs came from hens from a cage farm with access to artificial light. The next part of the eggs came from a farm with free-range walking in accordance with Art. 4 of Council Directive 1999/74/EC 4. The third part of the eggs came from an organic farm, ensuring strictly defined animal welfare (higher than conventional breeding) considering the maintenance conditions and proper manure management. The ecological farm was provided with the right number of hens, appropriate bedding (straw, shavings, 1/3 of the solid surface), feeders and drinkers (with permanent access), room microclimate, ventilation, natural and artificial lighting (15–20 lx for 16 h), 8-hour night rest (without lighting) and runs. The method of feeding hens in the cage system and from free-range was based on the same feed containing: corn, wheat, triticale, soybean meal, sunflower meal, fodder chalk, NaCl, 1-Ca phosphate, DL-Methionine, L-Lysine, Phytase, NEU-SOL, 0.5% premix, sodium bicarbonate, grindazim, aromabiotic. However, free-range hens had access to the enclosure (following Commission Regulation (EC) No 589/2008, Art. 4 of Council Directive 1999/74/EC 4), which was mostly covered with vegetation. Organic hens were fed with ecological mixtures of cereals, including wheat, triticale, corn, and organic root crops: potatoes, beetroots, and fodder carrots. Nettle and herbs, dried small-seed legumes, oilseeds, linseed up to 10%, and broad bean lupins were also used in the nutrition of laying hens. In addition, the laying hens also had access to ecological forage and the possibility of digging in the ground, where they met the demand for many needed minerals.

From each farm, 50 eggs were taken for further testing and comparative analysis of UPE. The average weights of the randomly selected eggs produced in the three systems were 56.7 g, 59.0 g, and 54.3 g, for caged, free-range, and organic eggs, respectively. The total number of eggs used in the analysis was 150.

2.2. UPE measurement

Measurements of emitted photons were made for intact eggs and separated yolks, albumen, and eggshells. The research was carried out using a proprietary measurement system, which enables the registration of photons emitted from the tested organic samples. The measurement equipment, presented in Fig. 1, works in an accredited laboratory under Polish Centre for Accreditation procedures (Polish Centre for Accreditation, 2020). The photon emission from eggs was measured using the HAMAMATSU type R4220 photomultiplier (Hamamatsu Photonics KK, Hamamatsu, Japan). It allows the analysis of electromagnetic waves with lengths ranging from 185 to 710 nm. The magnitude of the electric voltage controlling the gain of the photomultiplier was adjustable in the range from 300 V to 1500 V, where the adjustment step was 1 V. For that purpose power supply unit of high voltage (PSU HV) and SDS HT 1400 voltage regulator were used. National Instruments' LabView2015 graphical programming environment connected with the NI myDAQ measurement board acquiring photon data was used to control the measurement system. Time-Correlated Single Photon Counting (TCSPC) technique (Crockett et al., 2022; Wu & Hsueh, 2022) was applied for the recording of infrared decay.

Before the measurement of UPE, the eggs were washed under running water and dried at room temperature and atmospheric pressure. Before the experiments, each sample was placed in lightproof packaging and stored at 7–10 ° C for 24 h to maintain the same lighting conditions for all eggs and reduce the differences in the degree of exposure to the light during storage. The preparation time, between taking the sample from the refrigeration chamber and depressurising the lightproof packaging, did not exceed 30 min. An egg sample was placed in a measuring chamber thermally stabilised during the tests. The measurement process was carried out in a thermally stabilised room that eliminates solar radiation. Individual photons were counted

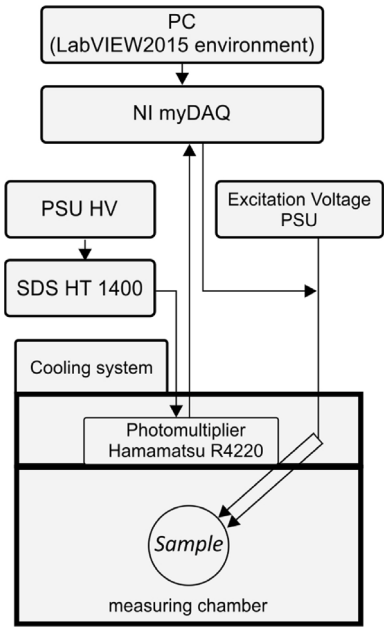


Fig. 1. Block diagram of author's photon emission measurement system based on a specially designed chamber with a photomultiplier detecting photons. LabVIEW 2015 — a graphical programming environment created by National Instruments, NI myDAQ — National Instruments measurement card of the photomultiplier signal, PSU HV — high voltage power supply unit, SDS HT 1400 — voltage regulator.

over an experimentally specified time interval to determine total ultra-weak photon emission. The minimum time for holding the sample in the lightproof chamber was assumed to be when the difference in the number of counted photons in two consecutive one-minute intervals was less than 10%. UPE was calculated as the absolute difference between the number of photons registered by the photomultiplier in the chamber with material (A) and the number of photons registered by the photomultiplier in the lightproof chamber without material (B), according to the formula L = A − B, where L is the number of photons emitted by the tested sample. Calibration of the sensors was carried out on each day of the measurements. Stabilisation of the system was performed to prevent disturbances resulting from temporary destabilisation of the standard conditions. This was the first phase of the measurement process, which lasted 120 s. Photon emission was then measured over time 600 s with a sampling rate of 4 Hz.

The ultra-weak luminescence measurement procedure was performed for whole eggs, and then, for the same samples, the measurement was performed for the separated egg components, such as white, yolk, and shell.

2.3. Data preprocessing

The dataframe contained $n_f = 4$ features represented by the numbers of biophotons emitted from $n_s = 150$ egg samples. The considered features were UPE from an intact egg and UPE from separated egg components — white, yolk, and shell. Initially, data outliers were removed by the capping method (Hodge & Austin, 2004; Mahmood, 2022), which uses sample mean and standard deviation values of a feature.

$$x'_i = x_i \left[(x_i > u_i - 3s_i) \wedge (x_i < u_i + 3s_i)\right], \quad i = 1, \ldots, n_f \tag{1}$$

where u_i and s_i denote the mean and standard deviation, respectively, of the values from the feature sample vector x_i. The lower and upper bound of the outliers were obtained by subtracting and adding three standard deviations to the mean as shown in Eq. (1). The whole sample row including the feature outlier was removed. The process was repeated for the sequence of feature columns until no more sample rows were deleted.

The data of each feature was independently centred and scaled according to the rule in Eq. (2) (Juszczak et al., 2002; Scikit-learn developers, 2022d).

$$x'_i = \frac{x_i - u_i}{s_i}, \; i = 1, \ldots, n_f \tag{2}$$

where x_i is the feature column vector, u_i and s_i are the sample mean and standard deviation, respectively, calculated in the set of n_s rows.

2.4. Egg classification

Different classifiers have been proposed for grading agricultural products including Support Vector Machine (SVM), Artificial Neural Network, Decision Tree, Random Forest, Bayesian Network, etc. In this paper, SVM was chosen for egg classification for its simplicity and based on the recommendation to use this technique for biological systems (Noble, 2006). It is also one of the most robust prediction methods based on the Vapnik–Chervonenkis theory (Cortes & Vapnik, 1995; Vapnik, 2000), which can perform both linear and nonlinear classification for the tested egg classes: caged, organic, and free-range. The robustness allows us to avoid overfitting. Moreover, SVM provides significant accuracy, is fast and is memory efficient for the tested dataset of several hundred items with four features, which can be considered rather small.

The decision function f of two class SVM classifier is obtained through the minimisation of the following expression:

$$\min_f \left[C \sum_{i=1}^{n_S} max\left(0, 1 - y_i f(x_i)\right) + \|f\|^2 \right], \tag{3}$$

where $\{x_i, y_i\}$, $i = 1, \ldots, n_S$, represents the dataset of photon emission samples $x_i \in \mathbb{R}$ and their associated class y_i, C is a regularisation parameter balancing the data fitting (left expression component) and the regularisation (right component). The multiclass classification problem is solved by decomposing it into multiple binary classification problems. In the original SVM version, the decision function is a hyperplane as in Eq. (4).

$$f(x) = w^T x - b, \tag{4}$$

where w is the plane orthogonal vector and b the vector constant. One of the strengths of SVM is its ability to choose a complex representation of the data thanks to using a kernel function that measures the similarity between samples. Eq. (5) shows the decision function using the Gaussian kernel $K(x, x')$ (also known as the Radial Basis Function (RBF)) providing a nonlinear separation surface between each class and the rest of them.

$$f(x) = \sum_{i=1}^{n_S} \alpha_i y_i K(x, x_i), \text{ where } K(x, x') = exp(-\gamma \|x - x'\|^2), \tag{5}$$

where $\alpha > 0$ is the Lagrange multiplier vector, $\|x - x'\|^2$ is the squared Euclidean distance between the two feature vectors, and γ denotes a parameter that controls the width of the Gaussian curve.

To check the tendency to clustering of egg samples based on measured photon emissions K-means method (Lloyd, 1982) has been applied. It relies on minimising the pairwise squared deviations of points in the same cluster as given in Eq. (6)

$$\arg\min_S \sum_{i=1}^{n_C} \frac{1}{|S_i|} \sum_{x,y \in S_i} \|x - y\|^2, \tag{6}$$

where x, y are different points from the same cluster $S_i \in S$, n_C is the assumed number of clusters. All types of discussed classifiers have been implemented in the Python environment. For the purpose of linear SVM

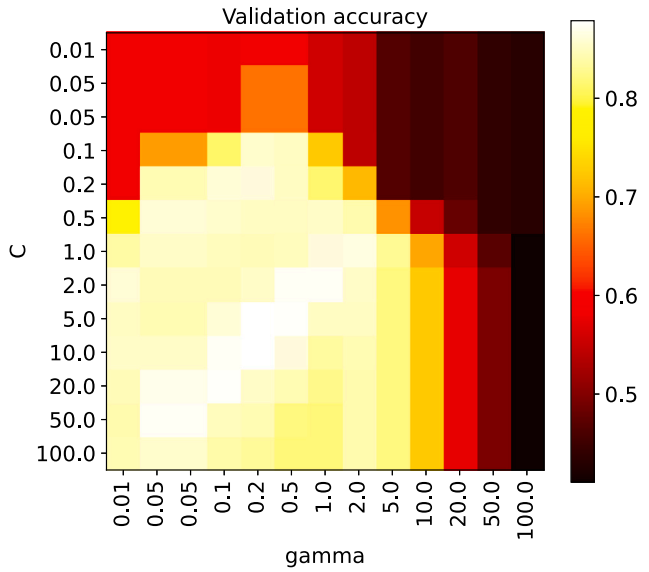


Fig. 2. Validation accuracy heatmap of SVM parameters (gamma, C) from Eq. (8) shown in quasi-logarithmic scale.

classification of egg types, the class *sklearn.svm.LinearSVC* has been used (Scikit-learn developers, 2022b). The class constructor shown in Eq. (7) has been called with 3 named parameters to fine-tune.

$$svc = LinearSVC(C, class\_weight, tol), \tag{7}$$

where svc denotes the classifier object, $C = 2$ is the regularisation parameter with value inversely proportional to the regularisation strength, the parameter *class_weight* = *'balanced'* adjusts class weights inversely proportional to their frequencies in the input data, $tol = 10^{-4}$ represents the tolerance for stopping criteria. In the case of RBF kernel (Eq. (5)) our SVM model has the parameters given in Eq. (8) (Scikit-learn developers, 2022c).

$$svc = SVC(kernel, C, class\_weight, gamma, tol), \tag{8}$$

where *kernel* = *'rbf'* denotes RBF type kernel, $gamma = 1$ is the kernel coefficient mentioned in Eq. (5), $C = 2$, $tol = 10^{-3}$ and *class_weight* = *'balanced'* parameters have the same meaning as in Eq. (7). The parameter values were optimised in a double quasi-logarithmic grid of $C \in [10^{-2}, 10^{2}]$ and $gamma \in [10^{-2}, 10^{2}]$ using 10-fold cross-validation functionality built in the *sklearn.model_selection.GridSearchCV* class. The heatmap in Fig. 2 illustrates the cross-validation accuracy encoded in colours. Among the several locations, with accuracy close to the maximum (≈ 0.9), the coordinates $C = 2$ and $gamma = 1$ were selected. Although the maximum accuracy for $C = 5$ and $gamma = 0.2$ is slightly higher than for $C = 2$ and $gamma = 1$, the authors prefer a variant of a simpler decision function with lower C and lower impact of a single training example expressed by a larger $gamma$ value.

To obtain K-means clustering of the tested eggs, the class *sklearn.cluster.KMeans* has been used, with the constructor in Eq. (9) (Scikit-learn developers, 2022a).

$$kmeans = KMeans(init, n\_clusters, max\_iter), \tag{9}$$

where *init* =*'random'* denotes initial centroids chosen randomly from the observation set, $n\_clusters = 3$ refers to the number of expected egg types, $max\_iter = 1000$ is the maximum number of algorithm iterations for a single run. The other method parameters retain their default values.

2.5. Performance analysis

To examine classification performance, two metrics based on the confusion matrix (Bhandari, 2020) have been applied: the overall accuracy (AC) (Eq. (10)) and Cohen's Kappa coefficient (K) (Eq. (11)) (Grandini et al., 2020). AC returns an overall measure of how much the model is correctly predicting on the entire set of data samples.

$$AC = \frac{c}{s}, \quad c = \sum_{i=1}^{n_C} C_{i,i}, \; s = \sum_{i=1}^{n_C}\sum_{i=1}^{n_C} C_{i,j} \tag{10}$$

where C represents the multiclass confusion matrix with elements $C_{i,j}$, n_C is a number of considered classes, c — the total number of elements correctly predicted, s — the total number of elements. Cohen's Kappa measures the concordance between predicted and true classes in the confusion matrix C, which are regarded as two random categorical variables (Cohen, 1960).

$$K = \frac{c \times s - \sum_{i=1}^{n_C} p_i \times t_i}{s^2 - \sum_{i=1}^{n_C} p_i \times t_i}, \qquad p_i = \sum_{k=1}^{n_C} C_{i,k}, \; t_i = \sum_{k=1}^{n_C} C_{k,i} \tag{11}$$

where p_i — the number of times that class k was predicted (column total), t_i — the number of times that class i truly occurs (row total). When $K = 0$ the model's prediction is totally independent of the true classification and if $K = 1$ the model's prediction is fully dependent on the actual classification. Instead, $K < 0$ means that the agreement between the predicted and the true classes distribution is even worse than the random agreement. The Z-test (Fleiss et al., 2003) was used to test the significance of Cohen's Kappa values.

Receiver Operating Characteristics (ROC) curves (Tharwat, 2018) were computed after SVM classification for each egg type to assess the dependence of the true and false positive rates on the position of the classifier cut-off point. The curves were obtained using the One vs. the Rest (OvR) method. The authors also carried out $F1$-score test of classification accuracy based on the *precision* and *recall* values (Grandini et al., 2020) for a single class of egg type compared with the rest of the classes. The $F1$ formula components are taken from the confusion matrix of classification results as in Eq. (12).

$$F1 = \frac{2}{precision^{-1} + recall^{-1}}, \tag{12}$$

2.6. Statistical analysis

The statistical significance of the differences between the emission of photons in the examined groups of eggs was determined using the one-way analysis of variance (ANOVA) (Brown & Forsythe, 1974). After the ANOVA analysis, the *post-hoc* analysis of the differences between subsequent pairs of housing types for a given UPE feature was investigated using the least significant difference (LSD) Fisher test. The Pearson correlation between UPE from different egg components was also investigated. Statistical analyses were performed in the Metabo-Analyst (https://www.metaboanalyst.ca) and PQStat programs (PQStat Software (2022). PQStat v.1.8.4. Poznan, Poland).

3. Results and discussion

Fig. 3(a) illustrates the average values of UPE from eggs and different parts of eggs recorded in total during 600 s. The number of photons emitted by each of the separately tested egg components was found to differ depending on the farming production system. For yolks, albumens, shells, and intact eggs, the highest ultra-weak emission was detected from ecologically farmed eggs. Primarily, the highest bioluminescence of the yolk and the white distinguishes this egg class from the examined egg population, where the average emission of biophotons over the entire measurement time was 119 and 122, respectively (Fig. 3(b)). For eggs obtained from free-range hens, the average photon emission from egg yolks and whites was 114 and 105, respectively, and for eggs from the cage system, these values were 104 and 105, respectively.

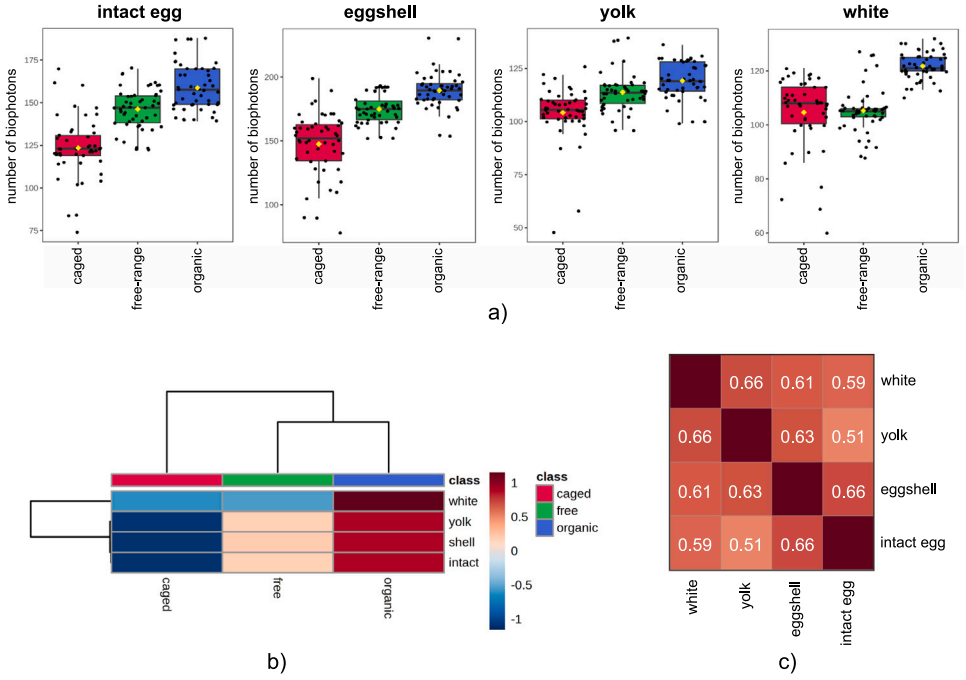


Fig. 3. UPE characteristic of egg components. (a) Mean values of UPE. The presented results show statistically significant differences for different types of hen rearing. For one-way ANOVA results, see Table 1. (b) Hierarchical clustering of egg UPE. (c) Correlation coefficient r of UPE from egg components.

Table 1
ANOVA test result of biophoton emission from eggs of various breeding types.

Egg part	F	p	Post-hoc test
Intact egg	73.62	<0.0001	free-range - cage; organic - cage; organic - free-range
Eggshell	70.28	<0.0001	free-range - cage; organic - cage; organic - free-range
White	51.84	<0.0001	organic - cage; organic - free-range
Yolk	25.80	<0.0001	free-range - cage; organic - cage; organic - free-range

The analysis of the data variance indicated that the UPE emission was significantly different for various rearing types of hens (Table 1). Therefore, the entire data set can be used as input features in the egg class identification process. However, we also tested various configurations of UPE feature vectors in which one chosen parameter was excluded. As can be seen in Table 1, UPE of egg yolks has the weakest ANOVA result (F-score = 25.8) among other features. In turn, as per *post-hoc* Fisher's test result, the UPE of egg white is not a discriminant between caged and free-range eggs. In this tested egg population, the total numbers of biophotons emitted by white and yolk are also quite well correlated with the correlation coefficient r = 0.66 (Fig. 3(c)); therefore, one of these features could be omitted. As a result, the following feature vectors of UPE from the egg and different egg components were considered:

- white, yolk, eggshell, and intact egg,
- white, eggshell and intact egg,
- yolk, eggshell, and intact egg.

Additionally, principal component analysis (PCA) was applied to reduce the dimensionality of feature space. We tested feature space reduced to the first two components in the classification process.

3.1. K-means grouping

K-means unsupervised algorithm, with defined k=3 centroids in the dataset, was tested to reveal the potential of splitting the tested egg population into natural groups of eggs regarding the ultra-weak photon emission. The method is prevalent, fast, and simple to implement for the known number of groups. It always guarantees convergence and allows improved classification accuracy by several restarts. Initially, the UPE feature vector contained 4 parameters: a total photon emission from white, yolk, eggshell, and intact egg, measured from eggs from three housing types. Table 2 and Fig. 4a present the distribution of egg samples in individual clusters. As can be seen, two distinguished clusters (cluster 1 and cluster 3) highly comply with the egg breeding types, organic and caged, respectively. However, in cluster 2, we observe the dominance of free-range eggs accompanied by eggs of other types, mainly those from cage breeding. As we previously noticed from Table 1, eggs from free-range breeding may be confused with cage eggs based on the white UPE. Therefore we may observe the most misclassifications in these groups. On the other hand, organically produced eggs are generally not confused with caged eggs, with only few exceptions. In this respect, organic eggs, which also had the highest emission of UPE, constituted the most homogeneous group. The compliance of the actual egg type with K-means grouping was checked using the Kappa Cohen test, based on the data in Table 2. The Kappa coefficient is statistically significant (p value < 0.000001) at the significance level of $\alpha = 0.05$, and amounts to $K = 0.61$, which proves a substantial agreement of the obtained clusters with the egg origin. As seen in Table 3, dataset pruning by removing sub-optimal features did not increase the accuracy of the K-means classification, except when data on biophotons emitted from whites were excluded from the input vector. The Kappa coefficient obtained for the UPE dataset without the UPE emission from the whites has a slightly better value ($K = 0.63$). In contrast, eliminating UPE data of egg yolks worsened the clustering result ($K = 0.42$). The result of egg classification carried out on the vector of the remaining 3 features (without UPE from whites) is presented in Fig.2b.

As shown, reducing the dataset by removing non-discriminatory feature slightly improved the result of grouping eggs according to their registered type. However, reducing features and orthogonalisation of data using PCA did not improve egg clustering. In the latter case, the K statistic of 0.46 for the loadings PC1 and PC2 indicates only moderate clustering agreement.

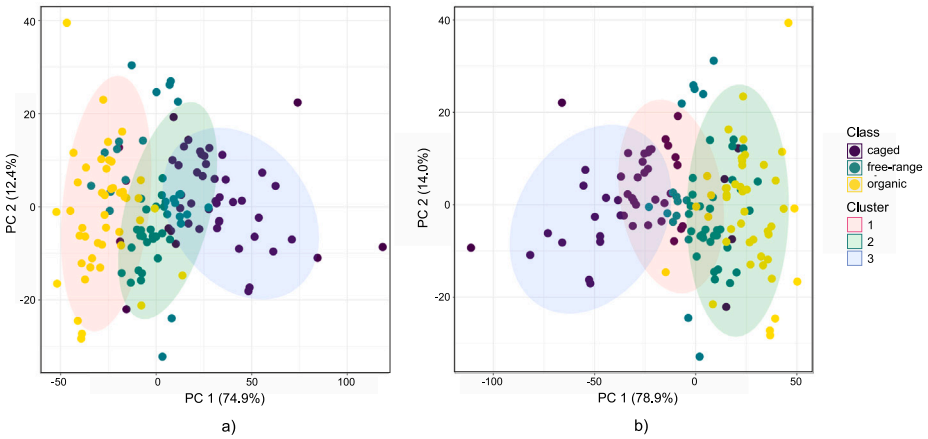


Fig. 4. Egg clustering by K-means using biophoton data. (a) Clusters based on 4 UPE features. For details, see Table 2; and (b) clusters based on 3 UPE features (UPE from yolks, shells, and intact eggs). For details see Table 5.

Table 2
Sample distribution after K-means algorithm clustering (4 feature input vector).

Egg type	Cluster 1	Cluster 2	Cluster 3
Organic	45	5	0
Free-range	9	41	0
Cage	4	20	24

Table 3
Cohen's Kappa test scores for K-means algorithm clustering results presented in Table 2, Table 4 and Table 5.

Emission components	K	Z-score	p
Yolk, white, eggshell, intact egg	0.61	10.87	<0.000001
Yolk, eggshell, intact egg	0.63	11.10	<0.000001
White, eggshell, intact egg	0.42	7.63	<0.000001
PC1, PC2	0.46	8.81	<0.000001

Table 4
Sample distribution after K-means algorithm clustering (3 feature input vector, without UPE from egg yolk).

Egg type	Cluster 1	Cluster 2	Cluster 3
Organic	42	8	0
Free-range	8	41	1
Cage	4	15	29

Table 5
Sample distribution after K-means algorithm clustering (3 feature input vector, without UPE from egg white).

Egg type	Cluster 1	Cluster 2	Cluster 3
Free-range	27	4	17
Organic	3	47	0
Cage	27	23	0

3.2. SVM classifiers performance

Two models of SVM with linear and radial kernel functions were developed to identify egg types. Confusion matrices verifying these models are presented in Table 6 and Table 7. They were calculated by validation tests on the dataframe randomly divided into training and test subsets, with the test size equal to 10% of $n_S = 150$ egg UPE data samples left after the outlier capping. Validation is performed on the egg-type prediction results of 1000 such tests with a random selection of test subsets. The 10% of test data was used instead of typical value 20% to keep more data of each class for training. With a limited number of egg samples per class, while remaining within the sample size limit for multivariate analysis, the 10:90 data split provides

Table 6
The confusion matrix obtained from the SVM classifier with the linear kernel function and performance evaluations, y — 'yolk', w — 'white', e — 'eggshell', ie — 'intact egg', 'f-range' — free-range, $F1$ — F1-score.

Emission sources	Predicted	True type					
		Cage	f-range	Organic	Precision	Recall	$F1$
y, w, e, ie	cage	3634	476	366	0.81	0.82	0.82
	f-range	366	4335	228	0.88	0.81	0.85
	organic	412	517	4666	0.83	0.89	0.86
w, e, ie	cage	3475	386	205	0.85	0.78	0.82
	f-range	642	4287	216	0.83	0.80	0.82
	organic	344	657	4788	0.83	0.92	0.87
y, e, ie	cage	3504	555	224	0.82	0.78	0.80
	f-range	623	3599	889	0.70	0.68	0.69
	organic	358	1117	4131	0.74	0.79	0.76

greater reliability for grading tests. The accuracy coefficients AC and K were calculated from the confusion matrices according to the rules in Eq. (10) and Eq. (11) and listed in Table 8 and Table 9, for linear and nonlinear classifier SVM, respectively. The classifier with RBF kernel function yields an accuracy AC of around 88%, which compares favourably with the accuracy of 84% of linear kernel SVM. Similarly, the Kappa coefficient decreases from 0.81 to 0.76 when replacing the RBF kernel with linear if one considers the photon emission from 4 egg components. In the Table 10 it has been shown that 20% of test data does not influence the classification accuracy.

The Receiver Operation Characteristic (ROC) curves of all egg types shown in Fig. 5 are computed based on raw output probabilities of the SVM predictions. Each ROC curve illustrates the dependency between *Sensitivity* and 1-*Specificity* of every egg type for all possible cut-off point values. Every egg class is compared with others in the mode One vs the Rest (OvR). The values of the Area Under the Curve (AUC) for each egg type have been added to the figure legends. In the case of linear SVM classification in Fig. 5(a), the ROC curve for organic eggs goes closest to the point (0, 1) and has the largest value $AUC = 0.96$ confirming the best discrimination of this egg type based on photon emission. The two other ROC curve profiles are further from the optimum trajectory and have lower AUC values, less than 0.91. The ROC curve of free-range eggs in Fig. 5(a) shows that this egg type with $AUC = 0.85$ may be more challenging to distinguish from the other types using linear classification. In the case of a non-linear classifier, the organic egg type has a cut-off point the same around (0, 1) as for other egg types. AUC values are similar for organic, free-range and caged eggs.

SVM classification accuracy of egg types depends on the amount and type of emission components introduced to the classifier. A full

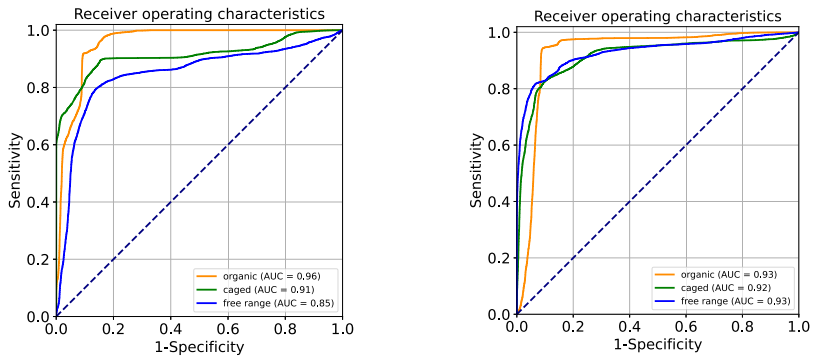


Fig. 5. ROC curves type One vs. the Rest (OvR) for all egg types recognised by the SVM classifier with the linear kernel (a) and RBF kernel (b), AUC — Area under the Curve. All emission components (yolk, white, shell, and intact egg) are used in the classification.

Table 7
The confusion matrix obtained from the SVM classifier with the radial kernel function and performance evaluations, y — 'yolk', w — 'white', e — 'eggshell', ie — 'intact egg', 'f-range' — free-range, $F1$ — F1-score.

Emission sources	Predicted	True type					
		Cage	f-range	Organic	Precision	Recall	$F1$
y, w, e, ie	cage	3695	387	139	0.88	0.83	0.85
	f-range	441	4469	193	0.88	0.86	0.87
	organic	337	368	4971	0.88	0.94	0.91
w, e, ie	cage	3475	386	205	0.86	0.80	0.83
	f-range	642	4287	216	0.85	0.81	0.83
	organic	344	657	4788	0.83	0.91	0.87
y, e, ie	cage	3499	297	89	0.90	0.79	0.84
	f-range	679	3982	868	0.72	0.77	0.75
	organic	276	877	4433	0.79	0.82	0.81

Table 8
SVM with the linear kernel function performance evaluation by Accuracy (AC) and Cohen's Kappa coefficient (K). p values of Z statistic < 0.000001.

Emission components	AC	K	Z-score
Yolk, white, eggshell, intact egg	0.84	0.76	132.03
White, eggshell, intact egg	0.84	0.75	130.43
Yolk, eggshell, intact egg	0.75	0.62	107.59
Intact egg	0.65	0.48	82.57
PC1, PC2	0.76	0.64	111.37

Table 9
SVM with the evaluation of radial kernel function performance by Accuracy (AC) and Cohen's Kappa score (K). p values of Z statistic < 0.000001.

Emission components	AC	K	Z-score
Yolk, white, eggshell, intact egg	0.88	0.81	140.53
White, eggshell, intact egg	0.84	0.76	130.43
Yolk, eggshell, intact egg	0.79	0.69	119.19
Intact egg	0.60	0.40	69.58
PC1, PC2	0.78	0.66	115.01

Table 10
SVM with the evaluation of radial kernel function performance by Accuracy (AC) and Cohen's Kappa score (K) for 20% of test data. p values of Z statistic < 0.000001.

Emission components	AC	K	Z-score
Yolk, white, eggshell, intact egg	0.87	0.80	193.09
White, eggshell, intact egg	0.84	0.76	182.77
Yolk, eggshell, intact egg	0.80	0.69	167.21
Intact egg	0.61	0.41	101.00

Table 11
SVM with the evaluation of polynomial and sigmoid SVM kernel functions performance by Accuracy (AC) and Cohen's Kappa score (K). p values of Z statistic < 0.000001.

Kernel	AC	K	Z-score
Polynomial, deg = 3	0.73	0.59	107.30
Polynomial, deg = 5	0.68	0.52	100.20
Sigmoid	0.74	0.62	106.02

set of 4 emission components, i.e. (yolk, white, shell, and intact egg) provides the best classification accuracy for the nonlinear SVM model, although they are correlated. Reducing the number of components by one lowers the K and AC coefficients to a varying degree, as illustrated in Table 8 and Table 9. The PCA transformation and reduction of input data to ($PC1$, $PC2$) components degrade the classifier accuracy to $AC = 76\%$ and $K = 0.64$ for linear kernel and $AC = 78\%$ and $K = 0.66$ for RBF kernel. This is because the number of emission data features of egg components is rather small; therefore, all features significantly influence the classification accuracy. ($PC1$, $PC2$) explain only about 87% of the data variance. Better classification accuracy with the RBF than with the linear kernel while maintaining the same level of C regularisation and moderate γ value proves that the considered problem of egg classification based on biophoton emission is not entirely linear. $F1$-score is a good quality indicator of an egg class identification relative to other classes in a multivariate environment. It balances both over and underestimation errors. The $F1$ values for different egg type classifications and different component sets emitting biophotons are given in Table 6 and Table 7. Relatively high values $F1 \geq 0.85$ for the RBF kernel SVM and all four or yolk, shell, and intact egg emission components confirm the advantage of non-linear classification in the studied cases. In the nonlinear case, organic eggs are best classified by $F1$-score, as shown for ROC curves.

In our experiments, the egg classification based only on the emission of intact eggs provides no more than 65% of accuracy (Table 8). Therefore, the authors propose to use biophoton emission measurement from the intact egg and its components' emission (yolk, white and eggshell). This approach allows an egg classification accuracy of almost 90% for the SVM with RBF kernel.

The results of UPE classification by SVM with kernels of sigmoidal and polynomial types have also been considered. The indices of classification accuracy of eggs were calculated for the UPE data from all egg components. As can be seen in Table 11, the accuracies for polynomial or sigmoidal kernels are worse even than in the case of linear kernels.

Our result of classification accuracy, which is 88% based on UPE measurement only, is located in the upper part of the precision range reported by other researchers, which is between 63–99%. Our accuracy does not differ from the quality of egg classification using different

Table 12
Egg quality parameters according to different researchers.

Quality params	Unit	Cage	f-range	Organic
Haugh unit[a]		84.60 ± 0.59	78.60 ± 0.36	–
Albumen pH[a]		8.11 ± 0.22	8.19 ± 0.38	–
Carotenoids[a]	μg/g	11.20 ± 0.20	39.10 ± 1.40	–
Tocopherol[a]	μg/g	91.30 ± 2.50	83.70 ± 1.70	–
Lutein[b]	μg/g	4.11 ± 3.75	7.49 ± 2.62	17.64 ± 4.30
Zeaxanthin[b]	μg/g	3.98 ± 1.38	3.25 ± 0.84	10.21 ± 1.99
Folic acid[c]	μg/100 g	78.50 ± 3.80	85.50 ± 5.70	113.8 ± 4.80
SFA[d]	%	36.26 ± 0.29	31.44 ± 0.38	–
MUFA[d]	%	46.90 ± 0.32	42.71 ± 0.80	–
PUFA[d]	%	16.84 ± 0.35	25.87 ± 0.76	–
n-6/n-3[d]	%	23.76 ± 0.95	14.48 ± 0.92	–
SFA[e]	%	84.45 ± nd	–	73.05 ± nd
MUFA[e]	%	108.45 ± nd	–	118.77 ± nd
PUFA[e]	%	35.85 ± nd	–	30.32 ± nd
n-6/n-3[e]	%	8.88 ± nd	–	4.76 ± nd
SFA[f]	%	34.90 ± 3.83	37.03 ± 7.09	33.63 ± 7.03
MUFA[f]	%	39.80 ± 3.72	40.95 ± 4.97	42.00 ± 8.76
PUFA[f]	%	25.30 ± 2.48	21.95 ± 3.57	24.38 ± 6.50
n-6/n-3[f]	%	10.85 ± 1.92	12.25 ± 3.88	11.53 ± 3.86

SFA — saturated fatty acids; MUFA — monounsaturated fatty acids; PUFA — polyunsaturated fatty acids; % of the fatty acids; nd — no data available.

[a] Gałązka-Czarnecka et al. (2019)
[b] Schlatterrer and Breithaupt (2006)
[c] Czarnowska-Kujawska et al. (2021)
[d] Popova et al. (2020)
[e] Mugnai et al. (2014)
[f] Egerer (2009)

models, including the SVM algorithm, based on NIR spectral data. Cruz-Tirado et al. (2021) built both PLS-DA and SVM models for Haugh Unit (HU) prediction, that reached accuracy of 87.0% and 85.7% respectively. In other works, even worse classification accuracy was achieved for egg freshness examination using NIR spectral technique. Zhao et al. (2010) developed classification models, including SVM, to discriminate fresh/unfresh eggs, achieving an accuracy of 63.3% of correct classification of fresh samples. Zhang et al. (2015) achieved freshness accuracy of 84% in egg freshness prediction.

The better accuracy of egg quality classification can be achieved by including a wide number of egg features. Sehirli and Arslan (2022) used over 20 egg features to predict egg freshness using different machine learning models. The accuracy of classification ranged from 81.0% (k-Nearest Neighbours - kNN model) to 98.6% (linear regression, LR). The accuracy of SVM model was also high, with its maximal value equal to 95.5%.

Lambing (1992) showed that the origin of hen's eggs could be determined based on the intact egg UPE intensity. Free-range eggs exhibited a higher emission rate after white light illumination because they have more stored energy UV than eggs from soil or cages. Köhler et al. (1991) found that exposure to sunlight or light similar to daylight, as well as feeding with green material, enhanced UPE from yolks of organic eggs. The quality of the chemical compounds of the eggs had a significant impact on the amount of internally stored energy.

Generally, high UPE values for eggs are obtained in conditions similar to the hens' natural living and feeding conditions. These, in turn, contribute to the high proportion of biologically active organic compounds in the eggs (Grashorn & Egerer, 2007; Lambing, 1992). In our experiment, the populations of the same hen breed of Rhode Island Red (R-11) were provided with various feeding methods and living conditions. In the case of ecologically farmed hens, a diet varied in grain and vegetables, with the addition of herbs, greens, and flax oil meal (rich in alpha-linolenic acid) was used. Chemical data for caged and free-range eggs from the same batch, as tested for photon emission, can be found in the work of Gałązka-Czarnecka et al. (2019). As you can see in Table 12, free-range eggs have a relatively high carotenoid content compared to caged eggs. Also, Table 12 presents selected egg quality indicators and their values for the considered types of hen farms, taken from various literature sources. Research by other authors shows that organic eggs also have a high proportion of carotenoids. According to data from Schlatterrer and Breithaupt (2006), organic eggs have more than twice the lutein and zeaxanthin content of other types of eggs. A lower proportion of compounds from the group of vitamins characterises cage eggs. In the related egg group, Gałązka-Czarnecka detected a lower proportion of tocopherol in cage eggs. In turn, Czarnowska-Kujawska et al. (2021) indicate the highest content of folic acid in the eggs from organic farming (113.8 μg/100 g), and the lowest in case of cage farming (78.5 μg/100 g).

The free-range housing is conducive to low values of n-6 polyunsaturated fatty acids (PUFA n-6) and a higher percentage of PUFA n-3 compared to eggs from a traditional cage farming system (Mugnai et al., 2014). These conclusions were also confirmed by Popova et al. (2020), Sergin et al. (2021) and Cartoni Mancinelli et al. (2022). It has been observed that the diet of hens kept with access to the run positively shapes the fatty acid profile of eggs. Hens' access to green pastures maintains a favourable, low ratio of n-6/n-3 acids in eggs.

According to Egerer (2009), yolk samples with low contents of saturated fatty acids (SFA) and high contents of PUFA showed the highest photon emissions. Organic eggs, containing the lowest SFA percentage in the whole fatty acid profile (see Table 12), were characterised by the highest emission of biophotons, equal to 2.38 qNL units. In her studies, Egerer found the lowest biophoton emissions from both caged and free-range eggs, around 1.8 qNL unit, which had a high content of SFA and low content of PUFA, respectively. The results of Grashorn and Egerer (2007) also depicted higher UPE with a slower declining trend for organic eggs. These were characterised by the lowest n-6/n-3 fatty acid ratios in yolks among free-range, barn and cage eggs, especially during summer. This confirms that the measurement of UPE can be a suitable method for evaluating the quality of organic eggs.

It is worth noting that, of the egg components, the largest emitter of photons is the shell separated from the whole egg. On average, this emission was higher than from yolks by 59%, 54%, and 42%, respectively, for organic, free-range, and cage eggs. This observation raises the interesting question of the dependence of photon emission on the structure of the shell and its mineral and biological components. This also suggests that in further UPE research, one may attempt a multivariate study, considering eggs' physical and chemical characteristics.

4. Conclusions

In the paper, we showed the possibility of distinguishing eggs from different breeding types based only on the differences in biophoton emission from individual parts of the eggs. We used UPE to discriminate three hen's egg categories: organic, free-range and caged. Using unsupervised K-means clustering, clusters corresponding to the tested types of eggs were confirmed in the space of features expressed by the emission of biophotons from individual egg components. Supervised SVM classifier training was performed based only on egg labelling. The nonlinear SVM classifier with RBF kernel allowed to achieve 88% accuracy of egg classification based on the amount of such emission. Although the emission counts of the different egg components are partially dependent, the use of the data of each one improves the final classification accuracy. The novelty of this work lies in building a classifier model based only on biophotons emitted by eggs, which differentiates them in terms of breeding types with the relatively high accuracy given above. This allows us to consider UPE as a carrier of important information, on par with spectral NIR data. So far, the authors have not found in the literature a UPE-based model that recognises the type of chicken egg breeding.

The results confirmed several previous reports that eggs from organic farming generally show higher UPE than eggs from other housing systems. They also highlighted the usefulness of analysing the number of emitted biophotons from whole eggs and their components to determine the type of hen housing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Joanna Sekulska-Nalewajko: Conceptualization, Formal analysis, Data curation, Visualization, Writing – original draft. **Jarosław Gocławski:** Conceptualization, Methodology, Validation, Software, Data curation, Writing – original draft. **Ewa Korzeniewska:** Conceptualization, Resources, Project administration, Writing – review & editing. **Paweł Kiełbasa:** Resources, Investigation, Writing – review & editing. **Tomasz Dróżdż:** Resources, Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Beloussov, L. (2003). Exploring the dynamic background of the developmental process and cell reactions with the use of an ultraweak photon emission. *BioSystems*, *68*, 199–212.

Bhandari, A. (2020). Everything you should know about confusion matrix for machine learning. https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/. (Last accessed 15 September 2022).

Brasil, Y., Cruz-Tirado, J., & Barbin, D. (2022). Fast online estimation of quail eggs freshness using portable NIR spectrometer and machine learning. *Food Control*, *131*, Article 108418. http://dx.doi.org/10.1016/j.foodcont.2021.108418.

Brown, M., & Forsythe, A. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics*, *30*, 719–724.

Cartoni Mancinelli, A., Mattioli, S., Twining, C., Dal Bosco, A., Donoghue, A., Arsi, K., Angelucci, E., Chiattelli, D., & Castellini, C. (2022). Poultry meat and eggs as an alternative source of n-3 long-chain polyunsaturated fatty acids for human nutrition. *Nutrients*, *14*, http://dx.doi.org/10.3390/nu14091969.

Cifra, M., & Pospišil, P. (2014). Ultra-weak photon emission from biological samples: Definition, mechanisms, properties, detection and applications. *Journal of Photochemistry and Photobiology B: Biology*, *139*, 2–10.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychological Measurement*, *10*, 37–48. http://dx.doi.org/10.1177/001316446002000104.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. http://dx.doi.org/10.1007/BF00994018.

Crockett, B., van Howe, J., Montaut, N., Morandotti, R., & Azaña, J. (2022). High-resolution time-correlated single-photon counting using electro-optic sampling. *Laser & Photonics Reviews*, http://dx.doi.org/10.1002/lpor.202100635.

Cruz-Tirado, J., Lucimar da Silva Medeiros, M., & Barbin, D. (2021). On-line monitoring of egg freshness using a portable NIR spectrometer in tandem with machine learning. *Journal of Food Engineering*, *306*, Article 110643. http://dx.doi.org/10.1016/j.jfoodeng.2021.110643.

Czarnowska-Kujawska, M., Draszanowska, A., Gujska, E., Klepacka, J., & Kasińska, M. (2021). Folate content and yolk color of hen eggs from different farming systems. *Molecules*, *26*, 1034. http://dx.doi.org/10.3390/molecules26041034.

Egerer, U. (2009). *Feldstudie zur Eignung der Biophotonenmessung für die Differenzierung von ökologisch und konventionell erzeugten Hühnereiern* (Phd thesis), Germany: Universität Hohenheim Stuttgart, Available at http://http://opus.uni-hohenheim.de/volltexte/2009/404/pdf/Diss.Egerer.pdf.

Fleiss, J., Levin, B., & Paik, M. (2003). *Statistical methods for rates and proportions* (3rd ed.). New Jersey: John Wiley & Sons, Inc., http://dx.doi.org/10.1002/0471445428.

Gałązka-Czarnecka, I., Korzeniewska, E., Czarnecki, A., Sójka, M., Kiełbasa, P., & Dróżdż, T. (2019). Evaluation of quality of eggs from hens kept in caged and free-range systems using traditional methods and ultra-weak luminescence. *Applied Sciences*, *9*(12), http://dx.doi.org/10.3390/app9122430.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. http://dx.doi.org/10.48550/ARXIV.2008.05756.

Grashorn, M., & Egerer, U. (2007). Integrated assessment of quality of chicken organic eggs by measurement of dark luminescence. *Polish Journal of Food and Nutrition Sciences*, *57*(4), 191–194.

Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*, 610–621. http://dx.doi.org/10.1109/TSMC.1973.4309314.

Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. *Intelligence Review*, *22*, 85–126. http://dx.doi.org/10.1007/s10462-004-4304-y.

Iida, T., Yoshiki, Y., Someya, S., & Okubo, K. (2002). Generation of reactive oxygen species and photon emission from a browned product. *Bioscience Biotechnology Biochemistry*, *66*, 1641–1645.

Juszczak, P., Tax, D., & Duin, R. (2002). Feature scaling in support vector data description. http://rduin.nl/papers/asci_02_occ.pdf. (Last accessed 15 September 2022).

Kemps, B., Bamelis, F., De Ketelaere, B., Mertens, K., Tona, K., Decuypere, E., & De Baerdemaeker, J. (2006). Visible transmission spectroscopy for the assessment of egg freshness. *Journal of the Science of Food and Agriculture*, *86*, 1399–1406.

Kobayashi, M. (2003). Spontaneous ultraweak photon emission of living organisms biophotons phenomena and detection techniques for extracting biological information. *Trends in Photochemistry & Photobiology*, *10*, 111–135.

Köhler, B. (2001). *Der Einfluss von Haltung. Fütterung und Beleuchtung auf die Biophotonenemission (delayed luminescence) sowie herkömmliche Qualitätsparameter von Hühnereiern*. Dipperz: Diss. Univ. Kassel und KWALIS Qualitätsforschung Fulda GmbH.

Köhler, B., Lambing, K., Neurohr, R., Nagl, W., Popp, F., & Wahler, H. (1991). 1. Photoenenemission: Eine neue methode zur erfassung der qualität von lebensmitteln. *Deutsche Lebensmittel-Rundschau*, *82*, 78–83.

Lambing, K. (1992). Biophoton measurement as a supplement to the conventional consideration of food quality. In *Recent advances in biophoton research and its applications* (pp. 393–414). Singapore: World Scientific Publishing.

Liu, Y., Ying, Y., Ouyang, A., & Li, Y. (2007). Measurement of internal quality in chicken eggs using visible transmittance spectroscopy technology. *Food Control*, *18*, 18–22. http://dx.doi.org/10.1016/j.foodcont.2005.07.011.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*.

Mahmood, S. (2022). Outlier detection (part 1): IQR, standard deviation, Z-score and modified Z-score. https://towardsdatascience.com/outlier-detection-part1-821d714524c. (Last accessed 15 September 2022).

McNamara, D. (2010). Eggs: A world of possibilities. *World Poultry*, *26*(7), 36–37.

Mehdizadeh, S., Minaei, S., Hancock, N., & Torshizi, M. (2014). An intelligent system for egg quality classification based on visible-infrared transmittance spectroscopy. *Information Processing Agriculture*, *1*, 105–114.

Meng, F., Chen, D., Li, X., Li, J., & Bao, J. (2014). Effects of large or small furnished cages on performance, welfare and egg quality of laying hens. *Animal Production Science*, *55*, 793–798.

Mugnai, C., Sossidou, E. N., Dal Bosco, A., Ruggeri, S., Mattioli, S., & Castellini, C. (2014). The effects of husbandry system on the grass intake and egg nutritive characteristics of laying hens. *Journal of the Science of Food and Agriculture*, *94*(3), 459–467. http://dx.doi.org/10.1002/jsfa.6269.

Nain, S., Renema, R., Korver, D., & Zuidhof, M. (2012). Characterization of the n-3 polyunsaturated fatty acid enrichment in laying hens fed an extruded flax enrichment source. *Poultry Science*, *91*, 1720–1732.

Noble, W. (2006). What is a support vector machine? *Nature biotechnology*, *24*, 1565–1567.

Nouri-Ahmadabadi, H., Omid, M., Mohtasebi, S., & Firouz, M. (2017). Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine. *Information Processing Agriculture*, *4*, 333–341.

Özbey, O., & Esen, F. (2007). The effects of different breeding systems on egg productivity and egg quality characteristics of rock partridges. *Poultry Science*, *86*, 782–785.

Polish Centre for Accreditation (2020). Testing laboratories. https://www.pca.gov.pl/en/accredited-organizations/accredited-organizations/testing-laboratories/AB%201698,entity.html. (Last accessed 15 September 2022).

Popova, T., Petkov, E., Ayasan, T., & Ignatova, M. (2020). Quality of eggs from layers reared under alternative and conventional system. *Brazilian Journal of Poultry Science*, *22*(1), http://dx.doi.org/10.1590/1806-9061-2019-1172.

Popp, F. (1992). Some essential questions of biophoton research and probable answers. In *Recent advances in biophoton research and its applications* (pp. 1–6). Singapore: World Scientific.

Popp, F. (2003). Biophotons - background, experimental results, theoretical approach and applications. In *Integrative biophysics* (pp. 387–438). Dordrecht: Springer.

Ruth, B., & Popp, F. (1976). Experimentelle Untersuchungen zur ultraschwachen Photonenemission biologischer Systeme. *Zeitschrift für Naturforschung*, *31*, 741–745.

Saifullah, S. (2020). Segmentation for embryonated egg images detection using the K-means algorithm in image processing. In *2020 fifth international conference on informatics and computing* (pp. 1–7). http://dx.doi.org/10.1109/ICIC50835.2020.9288648.

Saifullah, S. (2021). K-means clustering for egg embryo's detection based-on statistical feature extraction approach of candling eggs image. *Sinergi*, *25*(1), http://dx.doi.org/10.22441/sinergi.2021.1.006.

Saifullah, S., & Drezewski, R. (2022). Non-destructive egg fertility detection in incubation using SVM classifier based on GLCM parameters. *Procedia Computer Science*, *207*, 3254–3263. http://dx.doi.org/10.1016/j.procs.2022.09.383, https://www.sciencedirect.com/science/article/pii/S187705092201273X.

Saifullah, S., Drezewski, R., Khaliduzzaman, A., Tolentino, L., & Ilyos, R. (2022). K-Means segmentation based-on Lab color space for embryo detection in incubated egg. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, *8*(2), 175–185. http://dx.doi.org/10.26555/jiteki.v8i2.23724.

Saifullah, S., & Suryotomo, A. (2021). Identification of chicken egg fertility using SVM classifier based on first-order statistical feature extraction. *ILKOM Jurnal Ilmiah*, *13*(3), 285–293. http://dx.doi.org/10.33096/ilkom.v13i3.937.285-293.

Schlatterer, J., & Breithaupt, D. (2006). Xanthophylls in commercial egg yolks: quantification and identification by HPLC and LC-(APCI)MS using a C30 phase. *Journal of Agricultural and Food Chemistry*, *54*(6), 2267–2273. http://dx.doi.org/10.1021/jf053204d.

Scikit-learn developers (2022a). sklearn.cluster.KMeans. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html. (Last accessed 15 September 2022).

Scikit-learn developers (2022b). sklearn.svm.LinearSVC. https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html. (Last accessed 15 September 2022).

Scikit-learn developers (2022c). sklearn.svm.SVC. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html. (Last accessed 15 September 2022).

Scikit-learn developers (2022d). StandardScaler. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. (Last accessed 15 September 2022).

Sehirli, E., & Arslan, K. (2022). An application for the classification of egg quality and haugh unit based on characteristic egg features using machine learning models. *Expert Systems with Applications*, *205*, Article 117692. http://dx.doi.org/10.1016/j.eswa.2022.117692.

Sergin, S., Goeden, T., Krusinski, L., Kesamneni, S., Ali, H., Bitler, C., Medina-Meza, I., & Fenton, J. (2021). Fatty acid and antioxidant composition of conventional compared to pastured eggs: Characterization of conjugated linoleic acid and branched chain fatty acid isomers in eggs. *ACS Food Science & Technology*, *1*, 260–267. http://dx.doi.org/10.3390/nu14091969.

Tharwat, A. (2018). Classification assessment methods: a detailed tutorial. *Applied Computing and Informatics*, *17*(1), 168–192.

Vapnik, V. (2000). *The nature of statistical learning theory*. New York, USA: Springer-Verlag.

Wu, T., & Hsueh, T. C. (2022). A high-resolution single-photon arrival-time measurement with self-antithetic variance reduction in quantum applications: Theoretical analysis and performance estimation. *IEEE Transactions on Quantum Engineering*, 1–15. http://dx.doi.org/10.1109/TQE.2022.3209211.

Yannakopoulos, A. (2007). Egg enrichment in omega-3 fatty acids. In *Bioactive egg compounds* (pp. 159–170). Berlin: Springer-Verlag.

Yao, K., Sun, J., Chen, C., Xu, M., Cao, Y., Zhou, X., Tian, Y., & Cheng, J. (2022). Visualization research of egg freshness based on hyperspectral imaging and binary competitive adaptive reweighted sampling. *Infrared Physics & Technology*, *127*, Article 104414. http://dx.doi.org/10.1016/j.infrared.2022.104414.

Yenice, G., Kaynar, O., Ileriturk, M., Hira, F., & Hayirli, A. (2016). Quality of eggs in different production systems. *Czech Journal of Food Sciences*, *34*(4), 370–376.

Zhang, W., Pan, L., Tu, S., Zhan, G., & Tu, K. (2015). Non-destructive internal quality assessment of eggs using a synthesis of hyperspectral imaging and multivariate analysis. *Journal of Food Engineering*, *157*, 41–48. http://dx.doi.org/10.1016/j.jfoodeng.2015.02.013.

Zhao, J., Lin, H., Chen, Q., Huang, X., Sun, Z., & Zhou, F. (2010). Identification of egg's freshness using NIR and support vector data description. *Journal of Food Engineering*, *98*, 408–414. http://dx.doi.org/10.1016/j.jfoodeng.2010.01.018.