Computers and Electronics in Agriculture 229 (2025) 109755

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

An adaptive expert-in-the-loop algorithm for flock-specific anomaly detection in laying hen production

Lara A. van Veen [a,b,*], Henry van den Brand [b], Anna C.M. van den Oever [a], Bas Kemp [b], Ali Youssef [b]

[a] Vencomatic Group, Meerheide 200 5521 DW, Eersel, the Netherlands
[b] Adaptation Physiology Group, Wageningen University and Research, PO Box 338 6708 WD, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Egg production curve
Adaptive anomaly detection
Expert-in-the-loop
Incremental one-class SVM
Laying hen

ABSTRACT

Laying hen egg production shows high day-to-day intra-flock and inter-flock variability due to environmental stressors and suboptimal welfare, which negatively impact egg production. While laying hen flocks maintain efficient egg production until 100 weeks of age, the detection of health and welfare issues becomes increasingly important to prevent long-term effects on production and consequently on farmer economics. The intrinsic non-stationarity of continuously streaming production data imposes challenges on anomaly detection, rendering the current solutions for anomaly detection unsuitable for flock- and farm-specific, adaptive, high-confidence anomaly detection. In this paper, we propose an adaptive expert-in-the-loop algorithm for early anomaly detection in daily laying hen egg production. The key point was to dynamically model flock-specific egg production curves, using incremental one-class support vector machines (OCSVM), and compare daily acquired production data to an expert-defined adaptive reference trajectory, while allowing incorporation of variables related to hen performance or environmental variables. Detected anomalies receive an anomaly score based on a predefined normalized score threshold. Expert feedback is asked in instances of low-confidence anomalies, to iteratively improve accuracy of the anomaly detection algorithm. The proposed model was trained and tested, using real flock and synthetic datasets. Incremental learning improved anomaly detection precision from 0.70 to 0.81 compared to the initial OCSVM model. Expert feedback further refined the balance between sensitivity and precision, with an F1-score of 0.93 with 13% of expert feedback, thereby lowering false alarm ratios, while improving anomaly detection capabilities. Although this algorithm focusses on egg production, it can be adapted to detect anomalies in other production features, such as egg weight.

1. Introduction

The global laying hen sector is moving from cage housing systems towards alternative cage-free systems (Gautron et al., 2021). While alternative housing systems provide hens with significant opportunities to express natural behaviours, they bring implications for both farm management and layer health and welfare. Relevant health and welfare consequences of cage-free systems, in comparison to cage housing, include decreased air quality (Rodenburg et al., 2005), and increased risk on bacterial, viral and parasitic infections (Bonnefous et al., 2022). Monitoring layer health and welfare is more difficult and labour-intensive in multi-tier aviary and free-range systems, which are characterized by large flocks and free movement of layers throughout the three-dimensional space (Rodenburg et al., 2005).

Egg production is one of the primary indicators that allows continuous monitoring of layer flock health, welfare and performance. Egg production numbers are typically registered daily on all commercial layer farms and are frequently evaluated by the farm manager (van Veen et al., 2023). Egg production curves not only serve as flock performance indicators, but also enable prediction of future egg records, thereby allowing production planning and economic decision making regarding the optimal flock lifespan (Long and Wilcox, 2011). Flocks are replaced due to declining production and increased variability in egg quality (Dunn et al., 2005). Laying hen farmers are expected to move towards an extended flock duration of 90 to 100 weeks as response to societal demands for a reduced environmental footprint (Traore and Doyon, 2023). Breeding companies are selecting for increased laying persistency, while remaining stable egg quality and a healthy layer throughout flock

* Corresponding author.
E-mail address: lara.vanveen@vencomaticgroup.com (L.A. van Veen).

https://doi.org/10.1016/j.compag.2024.109755
Received 28 May 2024; Received in revised form 28 November 2024; Accepted 1 December 2024
Available online 11 December 2024

production (Arulnathan et al., 2024). Lengthening the flock duration emphasizes the need to detect health and welfare problems early to adapt management or feeding and to prevent long-term effects on egg production, thereby retaining profitable revenues (Traore and Doyon, 2023).

Because of the intensive nature of poultry production and relatively low profit margins, automatic monitoring and data-driven management are needed. A popular real-time monitoring and management concept in agriculture is Precision Livestock Farming (PLF), where algorithms and interfaces produce and visualize relevant information based on continuous, automated and real-time sensor data (Berckmans, 2017). PLF technologies in poultry production provide continuous daily data on feed intake and water intake, house temperature and relative humidity, egg numbers, weight and other external egg characteristics, and animal weight. On commercial farms, this data is mostly accessible in writing, but increasingly available as streaming data to the farm manager and archived on several hardware platforms.

Although streaming data is becoming available electronically, it is only used for reactive management by the farmer as opposed to predictive maintenance of performance, using algorithms and interfaces. The human expert, i.e. the laying hen farmer or veterinarian, is required to continuously interpret data allowing daily or weekly management decision-making, and failure to analyse the data efficiently renders them meaningless (Ribeiro et al., 2019). Challenges related to interpretation of streaming data for proactive maintenance include the absence of labelled malfunctions or health-related anomalies and the non-stationarity of the data. The cause of variation in egg production is multifactorial and characterized by an intrinsic non-stationarity or concept drifting (Kloska et al., 2023). The flock-specific shape of the production curve reflects natural variability, including breed differences and flock-dependent resilience against stress (Bedere et al., 2022), seasonal and daily variability due to for example environmental temperature (Li et al., 2020), (subclinical) diseases (Roberts et al., 2011), and welfare-related problems (Yamak and Sarica, 2012). Other sources of variability include changes in management and technical faults. With several potential explaining factors involved, expert knowledge of each individual laying hen flock is essential for interpreting the causes of production changes and for implementing appropriate measures to mitigate production losses.

Malfunctions and suboptimal performance in laying hen farms can be indicated as anomalies in the production curve. Traditional production-based anomaly detection relies on comparing commercial egg production data with static mathematical models (Grossman et al., 2000, Narinc et al., 2019). While these models provide a baseline for optimal production, they exhibit poor alignment with flock-specific and farm-specific production data. Additionally, they show limited robustness against dynamic influences by factors, such as age and environmental conditions. Few studies have yet addressed these limitations by exploring adaptive methods for real-time anomaly detection in egg production. Woudenberg et al. (2014) introduced an adaptive approach that effectively reduces the impact of natural variability by continuously adjusting detection parameters. However, such methods have primarily been validated on single flocks, raising concerns about their scalability and generalizability. Fuzzy logic has been applied to handle non-linear and imprecise egg production data (Omomule, Ajayi and Orogun, 2020). Emerging approaches in egg production modelling and anomaly detection exploit machine learning models, including Random Forest Modelling (Gonzalez-Mora et al., 2022) and artificial neural networks (Ramírez-Morales et al., 2017), and support vector machines (Ramírez-Morales et al., 2016), with prediction accuracies of 0.99. Machine learning methods offer promising advancements over traditional models, demonstrating higher accuracy in anomaly detection (Bumanis et al., 2023).

One significant challenge of machine learning application in poultry production lies in the sensitivity of machine learning algorithms to initial parameter settings, which can hinder its performance and robustness (Schmidl et al., 2022). Furthermore, the lack of adaptability in adjusting detection thresholds over the course of a flock round poses a significant limitation. Unsupervised learning methods, while efficient, often lack expert validation, leading to decreased detection accuracy (Ramírez-Morales et al., 2017). Expert knowledge is crucial for distinguishing between temporary anomalies and gradual changes in production patterns (Kloska et al., 2023). Supervised learning, on the other hand, is complicated by the low frequency of unexpected events in time series analysis (Schmidl et al., 2022), exacerbated by the variable timing of daily egg collection. Manual labelling of production defects based on egg production curves is impractical due to this variability. Therefore, recurrent expert input becomes essential for accurately identifying production anomalies amidst potential management irregularities.

This study aimed to introduce an algorithm for early detection of problems in laying hen egg production, which: 1) allows flock-specific comparison of actual egg data to an adaptive reference production curve; 2) detects anomalies in egg data, which is inherently subjected to fluctuations, with an unsupervised machines model that continuously learns from streaming data; 3) engages the expert, i.e. laying hen farmer and veterinarian, in several stages; and 4) allows for adding daily variables, such as climate data and hen-specific data, as extra features to improve model performance.

2. Data sets

2.1. Real-flock dataset

Real-flock data sets of 8 flocks of laying hens were obtained between August 2020 and June 2023. Laying hens from different breeds were housed on 4 commercial laying hen farms situated in the Netherlands, with aviary systems from an age of 17 or 18 weeks onwards (Table 1).

Duration of data collection varied between flocks from 453 to 612 days depending on data availability at start of the study. Variables of environmental conditions and hen performance were systematically obtained from an online farm management platform, which was connected to on-site farm computers. Variables were averaged per day to provide a comprehensive dataset for analysis. Environmental variables included the indoor poultry house temperature and indoor air pressure, and the outdoor temperature, which were expressed as daily averages based on a 15-minute interval measurement. These temperatures and pressure were measured by several thermometers and air pressure gauges, respectively. Temperatures were expressed as minimum, average and maximum indoor temperature and as minimum, average and maximum outdoor temperature. Air pressure was averaged.

Variables of hen performance included animal weight, measured by an automatic weighing scale that was present inside the laying hen house and determined daily average individual hen weight in grams based on voluntary weighing. Feed intake was expressed as total grams of feed per hen per day, based on summation of hourly feed weighing and division by the number of hens housed at start of the flock round. Water intake was expressed as total milliliters of water per hen, which was registered by a water meter inside the house. Daily egg production numbers were automatically counted by sensors each day in the afternoon. Egg production percentage was calculated by dividing the total number of eggs produced per day divided by the total number of hens that was housed in the facility at the start of flock production. Egg weight was determined based on the volume of eggs, measured by vision sensors (Meggsius Select, Vencomatic Group), multiplied by a density factor. Egg weight was determined for every egg on the egg belt and averaged per day.

The number of eggs with defects in or on the egg shell were measured with the vision sensors and expressed as number of defect eggs divided over the total number of counted eggs per day. Defects included blood stains, manure stains, yolk stains, bruises, breakages, and feathers. Individual eggs could receive more than 1 label.

Each of the 8 flocks was labelled as normal or abnormal based on

Table 1
Data overview of 8 laying hen flocks included in the study. Data points (n) are the number of days with production records available at the start of the study, including days (%) with a missing value for egg production.

Flock	Location	Breed	Nb. of hens housed	Data points (n) in days / % missing values	Hen age at placement in weeks	Normal/abnormal flock*
1	Farm A	Dekalb White	46,360	453 / 26%	17	Abnormal
2	Farm A	Dekalb White	46,360	510 / 4%	17	Normal
3	Farm A	Dekalb White	76,939	481 /26%	17	Abnormal
4	Farm A	Dekalb White	76,939	512 / 4%	18	Normal
5	Farm B	Lohmann LSL Lite	35,217	612 / 3%	18	Abnormal
6	Farm B	Lohmann LSL Classic	35,218	612 / 4%	17	Normal
7	Farm C	Lohmann Brown Classic	43,156	499 / 4%	17	Abnormal
8	Farm D	Lohmann Brown Classic	39,490	491 / 35%	17	Abnormal

* Based on visual inspection of the egg production curve. Egg production curves of normal flocks approximated trajectories from Grossman et al. (2000). Abnormal egg production curves had clear visual deviations from this model for longer periods than 1 day or daily outliers occurring with a continuously high frequency throughout the flock production period.

visual inspection of the egg production curve (Table 1). Normal egg production curves approximately followed the trajectory resulting from Grossman's egg production persistency model (Grossman et al., 2000), ignoring infrequently occurring outliers that only last 1 day. Abnormal egg production curves had clear visual deviations from this model for longer periods than 1 day, or daily outliers occurring with a continuously high frequency throughout the flock production period.

2.2. Synthetic dataset

In our proposed algorithm, an one-class support vector machines (OCSVM) model was solely trained on data belonging to normal flocks to detect instances that exhibit significant deviations from the defined normal behaviour and natural concept drifting in the egg production data. In the context of the present study, the available real-flock dataset contained limited number of normal production curves, specifically just 3 out of 8. Therefore, to support the training process of the initial OCSVM model, a synthetic dataset was generated based on the defined normal dataset using a Monte Carlo simulation (Ahmad, 2011). This simulation was conducted within the statistical characteristics of the original defined normal egg production time series, ensuring that the synthetic data closely mirrors the accepted normal behaviour. This augmentation is crucial to enhancing the model's training and overall performance. In theory, by introducing more synthetic training data points the model could be trained on diverse range of 'normal' production curves that could occur on commercial farms. Thereby, the model will be able to learn comprehensively the normal behaviour of the egg production process including the potential nonstationary variations (concept drifts).

2.2.1. Defining normal egg production curve

Egg production begins when the hens reach approximately an age of 18–22 weeks. A typical normal egg productivity curve of a high-productive flock (Fig. 1) is characterized by an initial phase of rapid increase in egg production, reaching its peak at around $P_{peak} = 95\text{–}97$ % (t_1). At the peak phase, the hens are laying eggs almost consistently at the maximum rate. The peak (plateau) lasts approximately 10 weeks, followed by a decline phase starting at approximately 38 weeks of age (t_2). At the decline phase, the production gradually diminishes, settling at approximately 90 % at 45 weeks after the start of egg production (Fialho and Ledur, 1997, Ahmad, 2011, Ramírez-Morales et al., 2016). The shape (i.e. the timing and height of the defined phases) of a normal egg production curve can vary, within an acceptable range, depending on several factors, such as breed, feed, management, and environmental variations.

The average and standard variations of egg production from the 3 normal egg production curves are depicted in Fig. 2. Weekly average egg production percentage was calculated for flock (2), flock (4) and flock (6), followed by taking the average and standard variation of egg production percentage across these flocks. The standard variations were

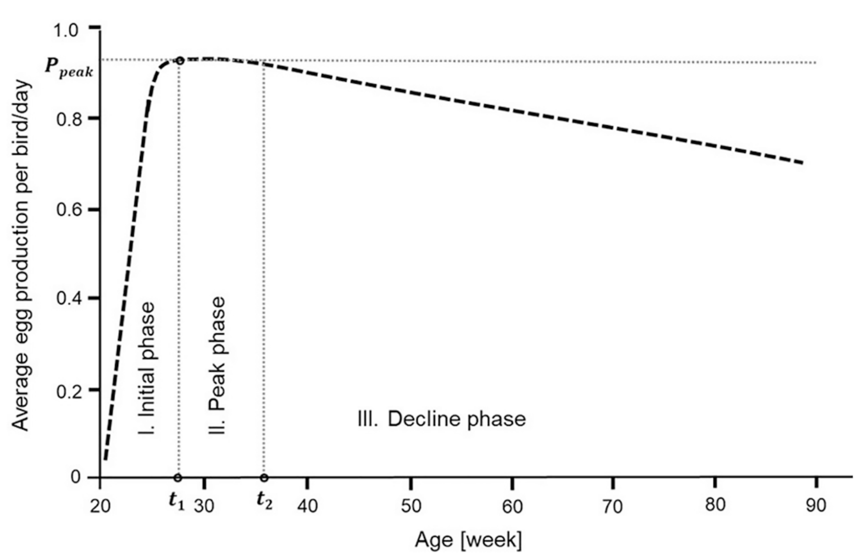

Fig. 1. A typical average egg production curve characterized by the initial phase, peak phase, and decline phase.

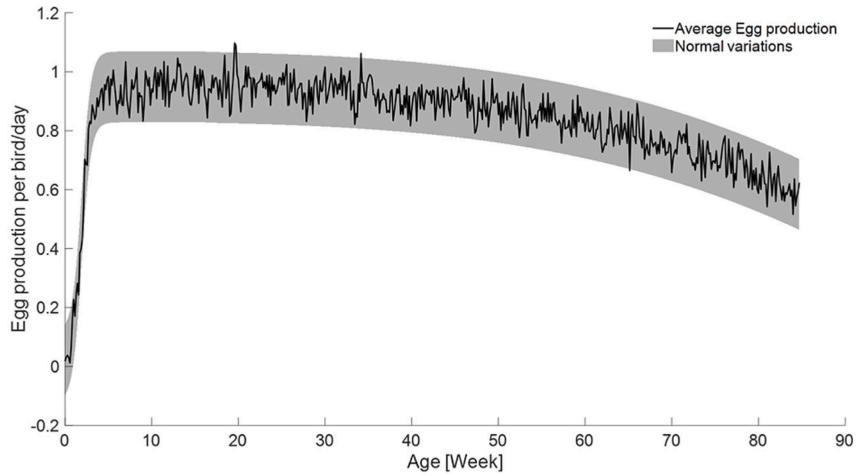

Fig. 2. Average egg production (eggs produced per hen housed at start) and standard deviations (normal variations) based on the 3 defined normal datasets.

considered in this context as acceptable normal variations and used in the Monte Carlo simulation.

2.2.2. Time series augmentation

The objective here was to create variations in the original time series data with feature-specific noise to generate N time series that capture the 'normal' variation patterns of egg production across all selected features. Let $\{(\boldsymbol{x}_i, y)\}$ be the original time series dataset, where $\boldsymbol{x}_i$ represents the i-th feature time series (e.g., egg production), and $y \in \{-1, 1\}$ is the binary label indicating normal (1) or anomaly (-1).

Transformation-based data augmentation: we used a jittering (Iwana and Uchida, 2021) function $\mathscr{J}(\boldsymbol{x}_i, y)$ that introduces feature-specific noise (Bishop, 1995, Iwana and Uchida, 2021) to the original data ($\boldsymbol{x}_i$, y) to generate N augmented time series data. The function $\mathscr{J}$ could be defined as follows: $\mathscr{J}(\boldsymbol{x}_i, y) = (\boldsymbol{x}_i', y)$, $i \in \{1, \cdots, m\}$, where $\boldsymbol{x}_i'$ is the augmented i-th feature with feature specific noise and m is the number of features. A feature-specific noise was introduced as follows:

$\boldsymbol{x}_{it}' = \boldsymbol{x}_{it} + \epsilon_{it}$, $t \in \{1, \cdots, L\}$, where $\boldsymbol{x}_{it}'$ is the t-th time step of the i-th feature, L is the number of data points of the time series, and $\epsilon_{it}\ \mathscr{N}(0, \sigma_i^2)$ is the Gaussian noise added at each time step t. The standard deviation σ is specific for each feature i.

Fig. 3 shows the synthetic time series data ($N = 100$) of the egg production feature showing acceptable normal variations within the egg production data.

3. Expert-in-the-loop incremental fault detection algorithm

The study introduces an innovative approach for fault detection in egg production systems by combining the strengths of machine learning and human expertise. Unlike other approaches (Grossman et al., 2000, Narinc et al., 2019), the proposed method, expert-in-the-loop incremental fault detection (EIFD) algorithm, employs an adaptive one-class support vector machines (OCSVM) model that dynamically adjusts to new data points, avoiding the reliance on a fixed model. The adaptive model operates incrementally, detecting anomalies in streaming data and ensuring accuracy through expert feedback. The expert-in-the-loop (EITL) approach is a form of an interactive machine learning (IML), in which there is a closer interaction between users (e.g., the laying hen farmer) and the learning systems, with experts interactively supplying information in a more frequent and incremental way compared to traditional machine learning (Amershi et al., 2022, Mosqueira-Rey et al., 2022). In this study, the expert is defined as the managing laying hen farmer, possibly supported by the poultry veterinarian affiliated with the farm. The proposed approach stands resilient against independently detected concept drift, providing a promising solution for real-time fault detection in dynamic egg production environments. The overall algorithm architecture is depicted in Fig. 4. The next paragraphs describe the main components of the EIFD algorithm pipeline.

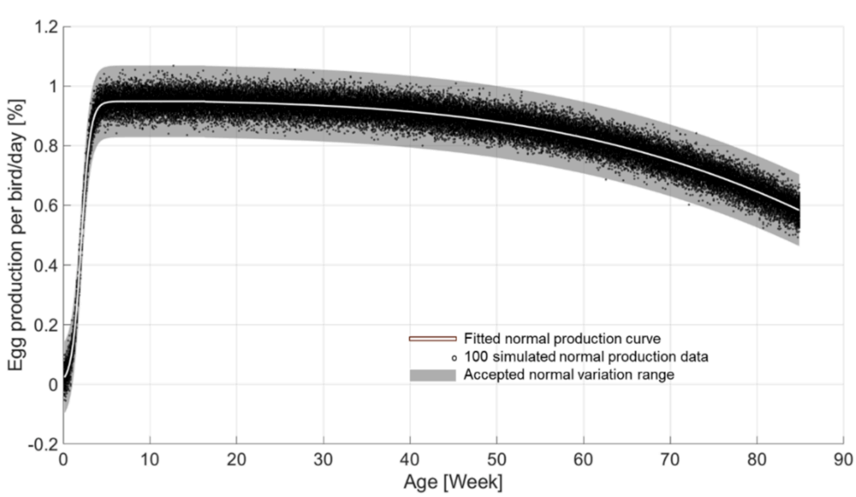

Fig. 3. 100 synthetic time series data of the egg production feature.

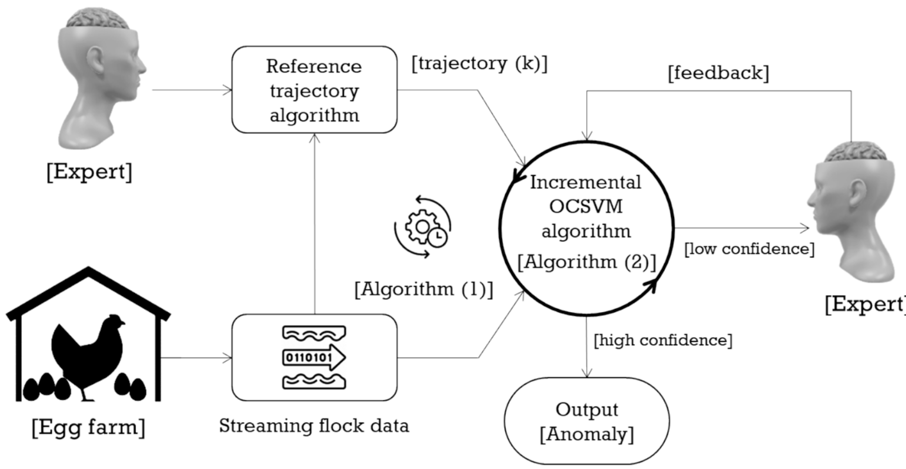

Fig. 4. Flow diagram illustrating the expert-in-the-loop incremental fault detection (EIFD) algorithm, depicting the incorporation of expert inputs to establish the reference (trajectory) egg production curve ([Algorithm 1]) and the feedback loop for labeling low-confidence detected anomalies (part of [Algorithm 2]), enhancing the incremental one-class support vector machines' anomaly detection capabilities.

3.1. Adaptive reference trajectory algorithm

For the best performance of the EIFD, a reference egg production curve is needed to provide a baseline or standard to which the actual egg production data can be compared. Initially, the baseline needs to be defined by the expert in such a way that it represent the expected trajectory of normal egg production under optimal conditions for a specific laying hen flock. However, due to the dynamic and drifting nature of the egg production process, the reference trajectory needs to be adapted continuously to fit the received streaming egg production data, otherwise any deviation from the initial trajectory will be considered an anomaly. Therefore, the mode parameters of the reference trajectory should be updated continuously. In this work, we introduce an adaptive reference trajectory (ART) algorithm to ensure the continual refinement of the trajectory model. The ART algorithm consists of the following:

3.1.1. Optimal trajectory model

From the available normal egg production curves, it was found that a piecewise logistic-quadratic model is the best fitting model to describe the data with an average goodness of fit ($R^2 = 0.99$). The piecewise logistic-quadratic (PLQ) model was mathematically expressed as a combination of logistic and quadractic functions. The reference trajectory curve is denoted by $\mathscr{R}(t)$, where t represent bird age. The PLQ model can be defined as follows:

$$\mathscr{R}(t) = \begin{cases} Logistic(t), \text{if } \ t < t_2 \\ Quadractic(t), \text{if } \ t \geq t_2 \end{cases} \tag{1}$$

where t_2 is the starting time of the decline phase (Fig. 1).

The logistic function ($\mathscr{L}$) was represented as:

$$\mathscr{L}(t) = \frac{P_{peak}}{1 + e^{-\kappa(t-t_1)}} \tag{2}$$

where P_{peak} is the peak (maximum) egg production, κ is the growth rate defining the slope of the egg production increases, and t_1 is the time at which the production reaches its peak value.

The quadratic function ($\mathscr{Q}$) was defined as follow:

$$\mathscr{Q}(t) = a(t - t_2)^2 + b(t - t_2) + c \tag{3}$$

where a, b, and c are the model parameters to be estimated from the observation data.

3.1.2. Initializing and updating the adaptive reference trajectory model

Firstly, the reference trajectory algorithm (RTA) need to be initialized by the expert. The expert, i.e. the laying hen farmer or veterinarian, should set the initial model parameters, namely, P_{peak} and t_1, for each flock separately. When no value is provided for P_{peak}, the setting is based on the stored reference production curves. These 2 parameters are enough to simulate the logistic part, $\mathscr{L}(t)$, of the reference trajectory model. By receiving more egg production data, the RT model needs to be updated, by fitting the PLQ equations (1) to the new data points. To ensure that there are enough data points to update the model parameters, a waiting period (in days), $\mathscr{W}_t \leq t_1$, is to be set. A waiting period of 10 days, equivalent to 10 samples, is found to be the minimum required to ensure that the adapted model parameters accurately reflect the dynamics of the received data. In this work, the model parameters of the PLQ equations (2) and (3) ere estimated, using the trust-region method (Coleman and Li, 1996) based on the nonlinear least-squares algorithm provided by the Matlab routine (LSQCURVEFIT). The proposed algorithm was summarized in the pseudo-code routine depicted in Algorithm 1.

Algorithm 1: updated reference trajectory

Initialization:
Set initial model parameters: P_{peak} and t_1 [to initiate $\mathscr{L}(t)$]
Set the waiting time period: $\mathscr{W}_t \leq t_1$ [default is 10 days]
Calculate the slope of the initial phase:$\kappa = \frac{P_{peak}}{t_1}$
Simulate the reference egg production, P_{ref}, using logistic part of the PLQ:
$P_{ref} = \mathscr{L}(t; P_{peak}, \kappa, t_1)$ for $0 \leq t \leq t_1$
Update the reference trajectory model
$t \leftarrow 0$ repeat
a. Wait until $t > t + \mathscr{W}_t$
b. Get the received egg production data $\{P(t) : 0 \rightarrow t + \mathscr{W}_t\}$
c. Exclude detected anomalies
d. Fit PLQ equations (2) and (3) to the received $P(t)$ data
e. Update model parameters; P_{peak}, t_1, t_2, a, b, and c based on the fitted PLQ equations (2) and (3)
f. Calculate κ (step 3)
Simulate the reference egg production P_{ref}
$\{P_{ref} : 0 \rightarrow t + \mathscr{W}_t\} = \begin{cases} L(t), \text{if } \ t < t_2 \\ Q(t), \text{if } \ t \geq t_2 \end{cases}$ $t \leftarrow t + 1$ until end of the input data stream

3.2. Incremental OCSVM (iOCSVM) algorithm

One of the essential characteristics of real-world dynamic time series, such as those related to egg production, is that the definition of abnormal or normal patterns tends to change over time (Ding et al., 2023), due to the changes in the statistical properties of the data. The phenomenon of such switching from pattern (i.e., concept) to another is known as concept drift. The concept at time t drift, according to number of publications, such as (Gama et al., 2014, Lu et al., 2018), is denoted as $\exists t : \mathrm{P}_t(\boldsymbol{x}, y) \neq \mathrm{P}_{t+1}(\boldsymbol{x}, y)$, or the change of joint probability $\mathrm{P}(\cdot)$ of $\boldsymbol{x}$ and y at time t.

Concept drift poses challenges to offline and batch-based anomaly detection algorithms. The dynamic nature of concept drift introduces complexities that can compromise the effectiveness of these algorithms to recognize and adapt to emerging patterns and changes in data distribution. Therefore, this paper proposes an incremental anomaly detection algorithm designed to effectively tackle the challenges posed by concept drift.

The incremental OCSVM algorithm (Schölkopf et al., 2001, Statnikov, 2011, Yokkampon et al., 2021), employed in our approach, is a form of online learning. Its practical advantage lies in its ability to incorporate additional training data as it becomes available, all without the need for complete re-training from scratch (Krawczyk and Woźniak, 2014). The main framework of the proposed iOCSVM algorithm is the pseudo-code routine (Algorithm 2) and the flowchart depicted in Fig. 5.

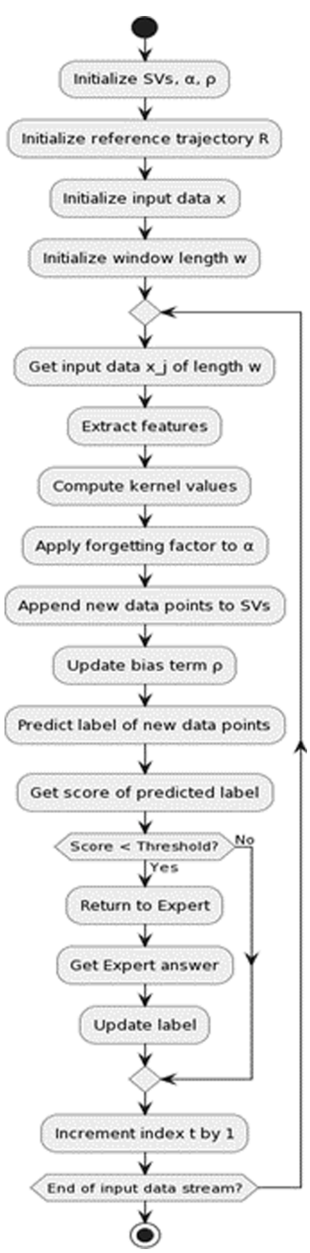

Fig. 5. Flowchart of the incremental one-class support vector machines (iOCSVM) algorithm for anomaly detection in egg production.

3.2.1. Initial OCSVM model

In this section, the formulation of the OCSVM is provided. Consider the training set $\{(\boldsymbol{x}_k, y_k)\}_{k=1}^{n}$ with input data $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \in \mathbb{R}^n$, where $n \in \mathbb{N}$ is the number of training samples, and output data $y_k \in \mathbb{R}$ with class labels $y_k \in \{-1, +1\}$. However, for the objective of the one-class SVM, the model is only trained on normal data with $y_k = +1$. The principle underlying OCSVM involves finding (learning) a hyperplane in the specified feature space that achieve a maximum separation between the training samples (i.e., representation of the normal class) from the origin (i.e., the only representation of the anomaly class) by maximizing the margin to the origin (Fig. 6) (Schölkopf et al., 2001, Statnikov, 2011, Yokkampon et al., 2021).

The OCSVM formulation uses a transformation function $\phi(\boldsymbol{x})$, defined by a kernel function (Schölkopf et al., 2001, Suykens et al., 2002, Statnikov, 2011), that maps the original feature space $\boldsymbol{x}$ into a higher dimensional feature space. The objective was to find a maximum margin to separate the training dataset from the origin by solving the following quadratic program (QP):

$$\min_{w,\xi,\rho} \left(\frac{\|w\|^2}{2} + \frac{1}{\upsilon n} \sum_{k=1}^{n} \xi_k - \rho \right) \tag{4}$$

subject to: $w^T \phi(\boldsymbol{x}) \geq \rho - \xi_k$, $\xi_k \geq 0$, where w is the weight vector, ρ is the offset term (or threshold), ξ_k is the slack variable for point k, n is the length of the training dataset, and $\upsilon \in (0, 1]$ is the regularization parameter.

When w and ρ solve the problem, then the decision function

$$f(\boldsymbol{x}) = \mathrm{sgn}[w^T \phi(\boldsymbol{x}) - \rho] \tag{5}$$

will return $+1$ for most of data points $\boldsymbol{x}$ in the training dataset (normal data points) and -1 otherwise (Schölkopf et al., 2001).

Schölkopf et al. (2001) solved the optimization problem in equation (4), using its dual problem formulation (for more information see Schölkopf et al. (2001). By driving the dual problem, the decision function, equation (2), can be transformed into a kernel expansion:

$$f(\boldsymbol{x}) = \mathrm{sgn}\left[\sum_{k}^{n} \alpha_k \mathrm{K}(\boldsymbol{x}_{\ell}, \boldsymbol{x}_k) - \rho \right] \tag{6}$$

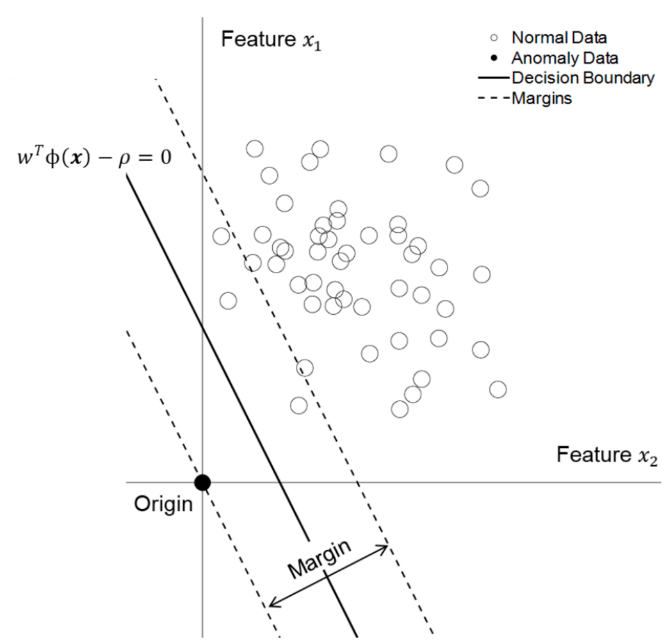

Fig. 6. Two-dimensional (2 features x_1 and x_2) example of the decision boundary learned by an one-class support vector machines (OCSVM) model.

where α_k is the Lagrangian multiplier, training samples $\boldsymbol{x}_k$ with corresponding nonzero α_k are called support vectors (SV), and $\mathrm{K}(\boldsymbol{x}_\ell, \boldsymbol{x}_k) = \phi(\boldsymbol{x}_\ell)^T \phi(\boldsymbol{x}_k)$ is a kernel function, such as the radial bias function (RBF) kernel

$$\mathrm{K}(\boldsymbol{x}_\ell, \boldsymbol{x}_k) = \exp\left(\frac{-\|\boldsymbol{x}_\ell - \boldsymbol{x}_k\|^2}{\gamma}\right) \text{for } \ell, k \in \{1, \cdots, n\}, \tag{7}$$

where γ is a hyperparameter that sets the width or the spread of the kernel.

During the training process of the OCSVM model, the number of hyperparameters are to be optimized including the regularization variable ν. By setting the ν parameter, we can control simultaneously the upper-bound on the fraction of the anomalies (faults) and the lower-bound on the fraction of the SVs.

3.2.2. Expert in the loop

In this proposed algorithm, we are giving the chance to the domain expert (e.g., laying hen farmer) to help enhancing the learning performance of the algorithm by labelling the observations that are considered to be anomalous by the algorithm or those which the algorithm is not confident with. In the context of anomaly detection, using the iOCSVM algorithm, each new data point is assigned a score (S_C) that indicate the degree of deviation from "normal" egg production dynamics. This score is derived based on the distance of each new data point to the decision boundary defined by the incrementally trained OCSVM model. These scores are used to rank data points in terms of their anomaly likelihood, allowing for threshold-based anomaly detection. While the threshold for anomaly detection is inherently defined within the iOCSVM algorithm, challenges arise in cases of concept drifting, where the predefined threshold may fail to effectively distinguish anomalies with sufficient confidence. In such scenarios, the expert can indirectly be involved in defining a new threshold by labelling instances of low-confidence anomalies based on a predefined normalized score threshold $S_{th} \in [0, 1]$. The magnitude of the S_{th} value indicates the extent to which expert input is required, with higher S_{th} value indicating a greater need for expert intervention in labelling anomalous instances. The expert intervention is measured by determining the percentage of the total test samples for which the expert was asked to provide labelling input.

3.2.3. Features and feature engineering

During the initial development phase of the algorithm, 4 input features were incorporated in the training process, which were the daily egg production, the daily feed intake, and the daily average bird weight. To enrich the model further and to find the optimal model, all available variables were iteratively added as extra features to study improvement of algorithm performance. As a result, 2 features, namely dynamic time-warping and feed-to-body weight ratio were crafted from the original inputs and included in the training process.

- Dynamic time-warping

To detect any deviation of the actual egg production (P) from the reference trajectory ($\mathscr{R}$) the Dynamic Time-Warping (DTW) similarity measure was used. DTW is commonly used to measure similarities between 2 time series (Aach and Church, 2001, Keogh and Ratanamahatana, 2005). In this case these are the reference trajectory $\mathscr{R} = \{r_1, r_2, \cdots, r_w\}$ and the actual egg production vector $P = \{p_1, p_2, \cdots, p_w\}$, $w \in \mathbb{N}$. Firstly, a distance or similarity function $\mathscr{D}(r_o, p_m)$ that measures the similarity between elements r_o and p_m of the sequences $\mathscr{R}$ and P, respectively, was calculated. In this work, the Euclidean distance is used as a similarity function. Then an accumulated distance matrix ($\mathscr{DM}$) of size $w \times w$, where $\mathscr{DM}(o, m)$ represents the cumulative cost of the alignment between elements r_o and p_m. Each element $\mathscr{DM}(o, m)$ is computed as follows:

$$\mathscr{DM}(o, m) = \mathscr{D}(r_o, p_m) + \min[\mathscr{DM}(o-1, m), \mathscr{DM}(o, m-1), \mathscr{DM}(o-1, m-1)] \tag{8}$$

The optimal warping path $Wp = \{wp_1, wp_2, \cdots, wp_r\}$ is defined as the path through the matrix $\mathscr{DM}$ with the lowest cumulative cost, where $r \in \mathbb{N}$ is the number of elements forming the optimal path. This is done by starting from element $\mathscr{DM}(0, 0)$ and recursively moving to adjacent cells with the lowest cost until reaching $\mathscr{DM}(w, w)$. Finally, the DTW distance between the sequences $\mathscr{R}$ and P is the normalized sum of costs along the optimal warping path as follows:

$$DTW = \sum_{i=1}^{r} Wp(i)/r \tag{9}$$

The DTW is implemented in a sliding window of size w as shown in Algorithm 2.

- Feed-to-body weight ratio

The feed-to-body weight ratio (FBW) is calculated as the ration between the daily feed intake (F_{in}) and the daily bird's weight (Wt_{Bird}).

$$FBW = \frac{F_{in}}{Wt_{Bird}} \tag{10}$$

Algorithm 2: incremental OCSVM algorithm

Inputs:
Trained initial OCSVM model parameters [SVs, α, ρ]
Reference trajectory $\mathscr{R}$ (Algorithm 1)
New data points with window length w
Outputs:
Updated OCSVM model parameters [SVs, α, ρ]
Labelled new data points $y_k \in \{-1, +1\}$
Initialization:
Let λ be the forgetting factor.
Set SVs be the set of the support vectors obtained from the initial training
Set α to the corresponding α values obtained from the initial training
Set the bias ρ to be the ρ obtained from the initial training
Initial reference trajectory $\mathscr{R}$ of length $\mathscr{W}_t$ (waiting period)
Initial input data $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{\mathscr{W}_t}\}$ of length $\mathscr{W}_t$
Incremental learning:
$t \leftarrow t + w$, where $w \le \mathscr{W}_t$ is the length of the sliding window
Repeat:
a. Get input data $\boldsymbol{x}_j$ of length w
b. Extract features according to equations (9) and (10)
c. Compute the kernel values $\mathrm{K}(\boldsymbol{x}_j, \mathrm{SVs})$ between the new data points and existing support vectors according to equation (7)
d. Apply the forgetting factor λ to the existing α values:$\alpha \leftarrow \lambda \times \alpha$
e. Append the new data points $\boldsymbol{x}_j$ to the SVs : $\mathrm{SVs} \leftarrow [\mathrm{SVs}\boldsymbol{x}_j]$
f. Update the bias term ρ based on the new SVs
Predict the label of the new data points according to equation (6) : $y_t \in \{-1, +1\}$
Get score of predicted labels S_C
If $S_C < S_{th}$ return to Expert (is this an Anomaly?)
Get Expert answer $y_{Ex} \in \{-1, +1\}$
$y_t \leftarrow y_{Ex}$ Increment index t by 1: $t \leftarrow t+1$ (sliding one time step)
Until end of the input data stream

4. Performance evaluation

For evaluation purposes, data was gathered from 4 abnormal egg production flocks, where each flock includes a different combination of normal and problematic instances (anomalies). In total, 1246 and 463 instances (i.e. days with digital egg production record) are labelled as normal and anomaly, respectively.

Evaluating the performance of anomaly detection algorithms differs from other classification algorithms primarily because the former is trained solely on 1 label (normal) dataset. Moreover, the testing dataset typically exhibits strongly imbalanced classes, a common scenario in anomaly detection problems. Thus, in this study, we combine accuracy metrics with the F1-score to provide a more comprehensive evaluation. Moreover, 2 additional criteria are employed: the total number of true anomalies detected, and the total number of false anomalies detected. These metrics provide the renowned trade-off between precision and sensitivity (recall) balance, offering deeper insights into the algorithm's performance.

4.1. Initial OCSVM model

The initial OCSVM model was trained, using the augmented dataset. The training dataset $\{(\boldsymbol{x}_k, y_k)\}_{k=1}^{n}$ comprised 100 augmented normal egg production data that included $n = 61300$ time instances, each labelled as normal (i.e., $y_k = +1$). Fig. 7 shows the distribution of resulting scores obtained during the training process. Approximately 90 % of the training data points were located within a mere 0.12 score difference (margin) from the score threshold of -0.88, as determined by the algorithm.

By testing the initial OCSVM model on the contaminated dataset, the resulting distribution of the scores, as depicted in Fig. 8, revealed the performance of the model. It is noticed that approximately 94% of the labeled anomalies are located on the right side of the score threshold. On the other hand, it is observed that about 80% of the labeled normal instances are situated on the left side of the score threshold, revealing a True Negative Ratio (TNR) of 80%. The performance evaluation results are summarized in Table 2. While the initial trained model exhibited a high sensitivity (recall) of approximately 0.94, the average precision value was approximately 0.70. This discrepancy underlines a significant trade-off: as the model demonstrates strong sensitivity in detecting anomalies, its precision suffers, leading to a non-negligible rate of false positives ratio (FPR) of approximately 14%. The occurrence of high sensitivity coupled with lower precision can be attributed to various factors, including inappropriate threshold selection, insufficient model tuning, and/or the presence of inherent concept drift within the data, as explained earlier. To address this issue, the proposed incremental algorithm (iOCSVM) incorporates an online learning technique that adaptively adjusts to changing (drifting) data patterns. Additionally, it employs a dynamic thresholding approach based on the score distribution of the incoming data streams, enabling the model to dynamically adapt its decision boundaries in response to evolving data dynamics.

4.2. Evaluation of the incremental algorithm (iOCSVM)

In this section, we compare the performance of the incremental algorithm (iOCSVM) under 2 conditions: with the incorporation of expert feedback to verify the labels of anomaly instances with low confidence scores, and without expert feedback.

4.2.1. Algorithm performance without the expert feedback

The performance evaluation results are summarized in Table 2. It is evident that the overall performance has enhanced in comparison to the initial OCSVM performance, as indicated by accuracy and F1-score metrics, which stands at 0.89 and 0.88, respectively. While the sensitivity of the iOCSVM algorithm did not exhibit significant improvement, there is a notable increase in precision to 0.81 compared to that of the initial OCSVM model, which had a value of 0.70. This resulted in a decline in the FPR value to 10%, compared to the 19% observed with the initial OCSVM model.

Fig. 9 shows the detected problematic egg production instances (anomalies) in egg production curve of flock (1) (Fig. 9A) and flock (5) (Fig. 9B), using the iOCSVM algorithm. It is shown that a decrease in egg production is identified as an anomaly before it becomes visually apparent.

4.2.2. Algorithm performance with the (simulated) expert feedback

As explained earlier (section 3.2.2), the extent to which the expert intervention is needed is determined by the magnitude of the predefined score threshold $S_{th} \in [0, 1]$. Thus, we have evaluated the performance of the iOCSVM algorithm at different values of S_{th}, indicating different levels of expert intervention as shown in Table 3.

In general, it is noticed that the overall performance of the algorithm is enhanced when expert feedback is introduced in the framework (Table 3) compared to a situation without expert feedback (Table 2). Anomaly detection accuracy improved with 4%-5% after incorporating the expert-in-the loop, while the F1-Score, which balances precision and sensitivity, improved with 8% at 5.2% of expert intervention compared to 0 % expert intervention. Referring to Table 3, the F1-Score shows an additional steady increase of 4 % as the score threshold rises from 5.2% to 13.0% The rising F1-Score indicates that the model achieves a better balance between correctly identifying anomalies and minimizing false positives as the threshold and level of expert intervention increase. This underscores the valuable role of expert feedback in improving precision and achieving a better balance between sensitivity and precision and consequently, refining the model's anomaly detection capabilities.

5. Discussion

5.1. Advantages of the proposed method

Despite the abundance of systematically collected data on farms, its full potential for laying hen health and welfare assessment, production planning and economic decision making remains largely unfulfilled. The implementation of anomaly detection algorithms in the context of on-farm streaming data presents a promising approach for optimizing

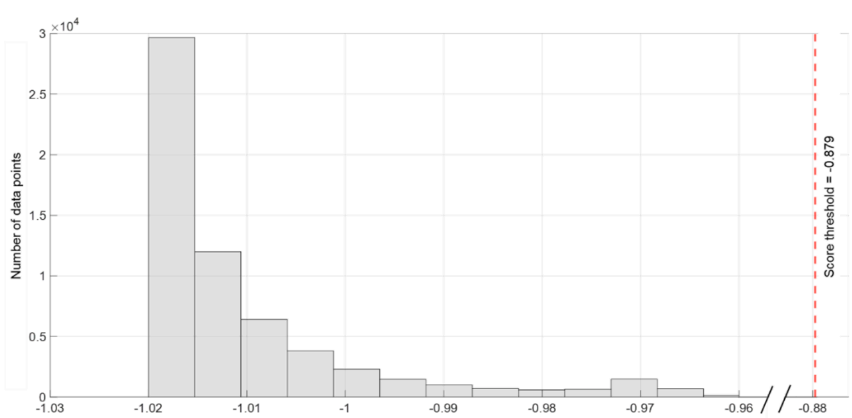

Fig. 7. The distribution of the resulting scores and the algorithm-determined score threshold during the training process.

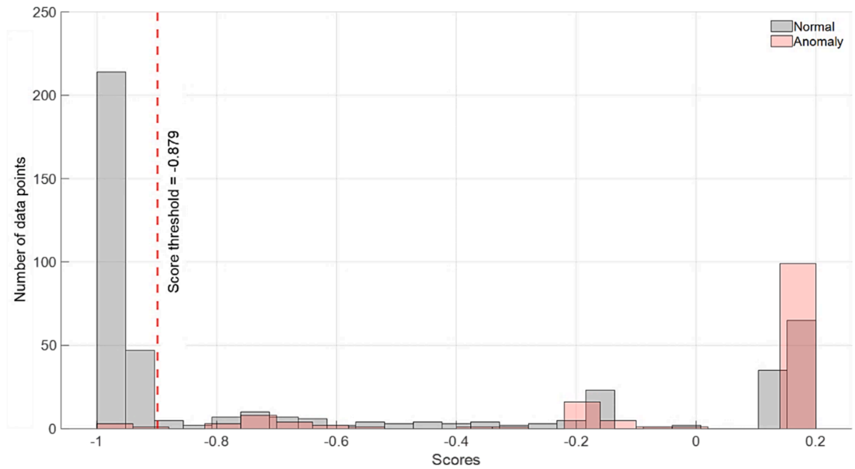

Fig. 8. The distribution of the resulting scores during the testing process.

Table 2
The performance evaluation results of the initial one-class support vector machines (OCSVM) model and the incremental one-class support vector machines (iOCSVM) model without expert feedback.

Model	FPR	Accuracy	Sensitivity	Precision	F1-Score
OCSVM	0.19	0.86	0.94	0.70	0.80
iOCSVM without expert feedback	0.10	0.91	0.96	0.81	0.88

daily decision-making processes in poultry production. This study makes a first step towards adaptive, flock-specific anomaly detection based on real-time egg production records, using the strength of expert-enhanced one-class support vector machines.

Central to our approach is the integration of the expert opinion, facilitating the wish of farmers to not only be alerted to anomalies, but also to be involved in the data underlying the alerts (Lokhorst and Lamaker, 1996). Therefore, our system empowers laying hen farmers and other experts, such as veterinarians, to view the data driving anomaly cases, enabling informed decision-making and subsequent diagnosis. A number of farms affiliated with this project currently receive real-time sensor-based egg counts through a digital platform, facilitating continuous monitoring and visualization of production metrics. Experts from these farms, such as laying hen farmers, veterinarians and feed advisors, could engage in the future in initial algorithm implementation and feed the incremental learning system via this platform.

Considering the evolving landscape of farming, younger farmers are likely to be willing to adopt and think with companies that aim to improve farming practices (Schukat and Heise, 2021). However, it's important not to overlook the expertise of older farmers, who may not be as connected to data-driven solutions, but still possess invaluable knowledge derived from years of experience. We aim to bridge the gap between traditional farming and modern technology using domain knowledge from laying hen farmers, preferably with at least 5 years of field experience into developing the modern anomaly detection algorithms, ensuring a smoother transition towards more efficient and sustainable farming practices. Continuous algorithm-based anomaly detection will become particularly important, especially within larger farms and among those with personnel characterized by scarcity and diverse educational backgrounds.

We emphasize the necessity of retaining expert involvement even after deploying and re-current training of the algorithm on commercial farms. The incremental learning model is specifically engineered to adapt and improve over time, learning from instances where confidence in flagged anomalies is low. False alarm ratios are higher without the expert-in-the-loop than with the expert-in-the-loop as apparent from precision scores, even when normalized score threshold is low. False alarms can hinder adoption of technologies and negatively impact farmer work satisfaction and human-animal relations and thereby laying hen welfare (Tuyttens et al., 2022). Experts can adjust the normalized score thresholds to determine what constitutes normal behavior of egg production in their respective flock, and thereby influence false alarm ratios. While they provide opinions on flagged low-confidence abnormalities, they also have the authority to remove normal flags when they deem them to be abnormal. With continuous streaming flock data, experts are inclined to recalibrate thresholds with improved algorithm performance.

The current approach optimizes daily hen health and welfare assessment by comparing daily egg production data to an adaptive optimal reference production to identify abnormal egg production instances. The initial values for the reference trajectory, which are the timing and height of peak production, are derived from a combination of farmer experience with farm and flock-specific performance, rearing background, and standard values based on breed guidelines, allowing for a tailored approach to anomaly detection. The flock-specific nature of the model may enhance its external validity by increasing its relevance and accuracy in real-world settings, which is one of the main threats posed by precision livestock farming to animal welfare (Tuyttens et al., 2022).

5.2. Limitations and future working points

Incremental model performance can be improved by adding other egg production-related variables than daily feed-to-body weight ratio. The current proposed algorithm aimed to allow integration of diverse environmental and physiological variables, such as environmental temperature and hen weight. The integration creates adaptability to the intrinsic concept drift present in laying hen production, for example in case of chronic stress by poultry red mite infestation with prolonged effects on farm productivity (Sigognault Flochlay, Thomas and Sparagano, 2017). It enhances functionality (Gumiran, 2024) and precision (Bumanis et al., 2023), while reducing the need to manually incorporate data features into the assessment. We foresee laying hen farmers having the autonomy to select the variables as features they deem relevant, thereby expanding the capacity for flock- and farm-specific anomaly detection. Multi-feature modeling not only saves time, but also enhances the objectivity of anomaly detection, improves the predictive power of egg production when features have a direct effect on egg output (Lokhorst and Lamaker, 1996, Yin et al., 2023) and ultimately improves hen health and welfare assessment.

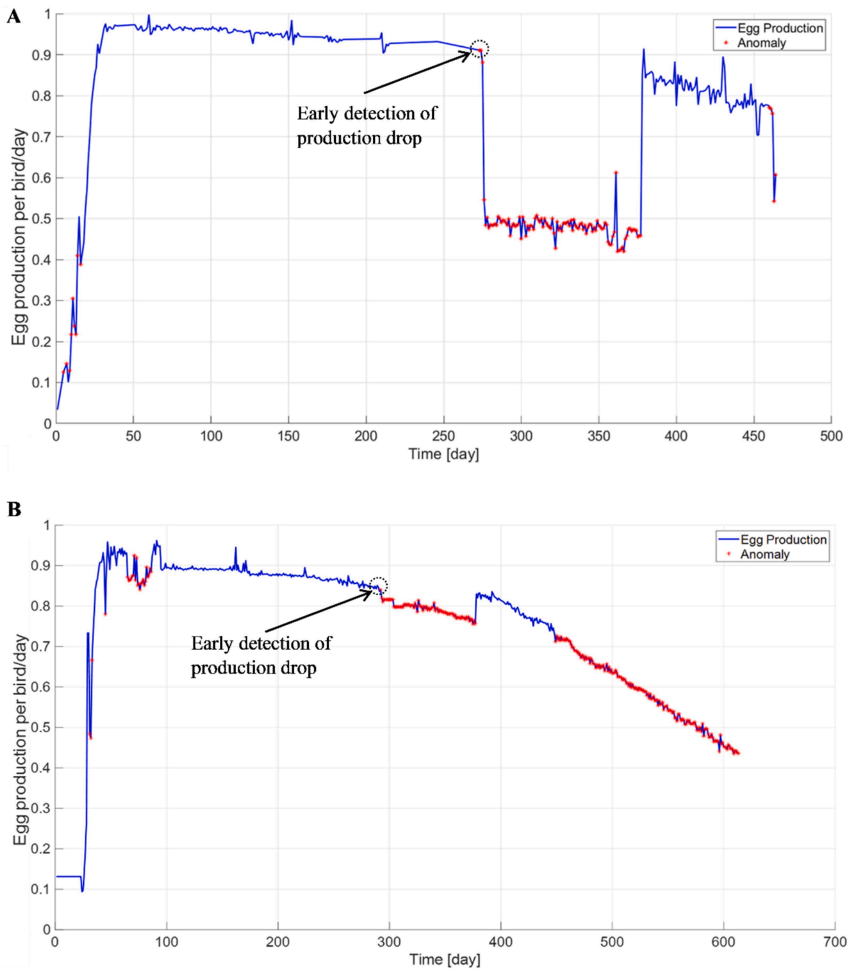

Fig. 9. The egg production curve of flock (1) (**A**) and flock (5) (**B**) combined with the anomalies detected (highlighted with red dots), using the incremental one-class support vector machines (iOCSVM) algorithm. Early detected drops in egg production by the algorithm are highlighted.

Table 3
The performance evaluation results of the incremental one-class support vector machines (iOCSVM) algorithm with different level of expert intervention simulated by employing different score threshold.

Score threshold S_{th}	Accuracy	Sensitivity	Precision	F1-Score	Expert intervention
0.35	0.95	0.94	0.85	0.89	5.2%
0.37	0.96	0.94	0.91	0.92	6.0%
0.39	0.96	0.94	0.92	0.93	7.6%
0.41	0.96	0.93	0.94	0.93	13.0%
0.43	0.96	0.93	0.94	0.93	13.4%
0.45	0.96	0.94	0.93	0.93	13.6%

We studied the effect of incorporating various external egg characteristics, such as egg shell shape and integrity, and shell soiling, into the current model on model performance. These traits, although infrequently recorded with on-farm sensors, may serve as indirect indicators of environmental conditions, nutritional status, or stress levels experienced by laying hens and possibly precede deviations in egg numbers, leading to even earlier detection of health problems in laying hens (Roberts et al., 2011, Yamak and Sarica, 2012). It is expected that in the near future, computer vision or other sensor techniques will increasingly facilitate monitoring of these characteristics in commercial farms. According to Lokhorst and Lamaker (1996), the natural temporal fluctuations observed in the number of second-grade eggs and floor eggs may however pose challenges in their utility for monitoring daily production processes in aviary laying hen systems.

The available egg characteristics were not included in the proposed model, because they did not improve anomaly detection performance, due to several reasons. These reasons include varying thresholds of labelling egg characteristics between farms, incomplete data sets for certain flocks, and a lack of optimization in labeled variables. Data quality characteristics, such as uniformity and completeness affect forecasting effectiveness of egg production (Bumanis et al., 2023). Despite the limitations in the current dataset, it's noteworthy that advancements in both sensors and sensor-data handling are rapidly progressing, continuously refining and enhancing data quality for more effective use in animal management practices (Carletto, 2021). This opens up promising opportunities for production forecasting. Determining the right number and timing of lagged or leading features, along with the suitable forecast window size, is crucial. It affects prediction complexity, accuracy, and practicality in daily management decisions (Bumanis et al., 2023).

There are several opportunities for prospective advancement of the proposed conceptual framework. One such opportunity involves the adaptation of the model to detect anomalies in daily streaming egg weight data besides egg numbers, provided it is consistently collected on laying hen farms. This adaptation could facilitate defect detection utilizing uniformity in egg weight as key variable (Ji et al., 2025). Egg weight holds promise as a potential early indicator of hen health issues, as it remains unaffected by the variable timing of egg collection by farmers. Another opportunity extends beyond poultry farming, with potential applications in anomaly detection within the lactation curve for dairy cows. These future directions show the versatility and potential for broader agricultural applications.

Furthermore, the implementation of our anomaly detection model holds significant implications for scientific knowledge advancement in the poultry domain. By establishing relationships between various parameters, such as age at initial egg production and persistency of lay, our model contributes to a deeper understanding of flock dynamics and performance trends (Gautron et al., 2021). Through the integration of flock-based metadata, our approach facilitates the comparison of flocks, even with variations in set-up times and individual laying hen characteristics.

6. Conclusions

Our approach on anomaly detection based on egg production data represents a significant step towards unlocking the full potential of on-farm data in optimizing flock management, health and welfare. By integrating expert insights with advanced, adaptive algorithms, we aim to detect and predict problems in poultry health and welfare objectively. The incremental one-class support vector machines model showed an anomaly detection accuracy of 0.96 after incorporation of expert intervention, an improvement of 10% compared to the initial OCSVM model and 5% compared to the incremental OCSVM model without expert intervention. The F1-Score reached 0.93 at 13% of expert intervention, as opposed to 0.81 at 0% of expert intervention. The proposed algorithm supports daily decision-making in poultry farming and improves domain knowledge on predictive flock-level indicators of laying hen health and welfare, thereby enhancing flock health, productivity, and overall farm profitability.

CRediT authorship contribution statement

Lara A. van Veen: Writing – review & editing, Writing – original draft, Visualization, Project administration, Investigation, Data curation, Conceptualization. **Henry van den Brand:** Writing – review & editing, Supervision, Conceptualization. **Anna C.M. van den Oever:** Writing – review & editing, Supervision, Data curation. **Bas Kemp:** Writing – review & editing, Supervision, Conceptualization. **Ali Youssef:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization.

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.

Acknowledgements

This research received financial support from Vencomatic Group, and Vencomatic Group collected the production data from 4 laying hen farmers in the Netherlands. The authors appreciate the permission of laying hen farmers to use their flock data for research purposes. The funding source had no involvement in study design, data analysis nor data interpretation.

Data availability

Data will be made available on request.

References

Aach, J., Church, G.M., 2001. Aligning gene expression time series with time warping algorithms. Bioinformatics 17 (6), 495–508. https://doi.org/10.1093/bioinformatics/17.6.495.

Ahmad, H.A., 2011. Egg production forecasting: Determining efficient modeling approaches. J. Appl. Poult. Res. 20, 463–473. https://doi.org/10.3382/japr.2010-00266.

Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T., 2022. Power to the people: The role of humans in interactive machine learning. AI Mag. 35 (4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513.

Arulnathan, V., Turner, I., Bamber, N., Ferdous, J., Grassauer, F., Doyon, M., Pelletier, N., 2024. A systematic review of potential productivity, egg quality, and animal welfare implications of extended lay cycles in commercial laying hens in Canada. Poult. Sci. 103 (4), 103475. https://doi.org/10.1016/j.psj.2024.103475.

Bedere, N., Berghof, T.V.L., Peeters, K., Pinard-Van der Laan, M.H., Visscher, J., David, I., Mulder, H.A., 2022. Using egg production longitudinal recording to study the genetic background of resilience in purebred and crossbred laying hens. Genet. Sel. Evol. 54, 26. https://doi.org/10.1186/s12711-022-00716-8.

Berckmans, D., 2017. General introduction to precision livestock farming. Anim. Front. 7 (1), 6–11. https://doi.org/10.2527/af.2017.0102.

Bishop, C.M., 1995. Training with noise is equivalent to Tikhonov regularization. Neural Comput. 7 (1), 108–116. https://doi.org/10.1162/neco.1995.7.1.108.

Bonnefous, C., Collin, A., Guilloteau, L.A., Guesdon, V., Filliat, C., Rehault-Godbert, S., Rodenburg, T.B., Tuyttens, F.A.M., Warin, L., Steenfeldt, S., Baldinger, L., Re, M., Ponzio, R., Zuliani, A., Venezia, P., Vare, M., Parrott, P., Walley, K., Niemi, J.K., Leterrier, C., 2022. Welfare issues and potential solutions for laying hens in free range and organic production systems: A review based on literature and interviews. Front. Vet. Sci. 9, 952922. https://doi.org/10.3389/fvets.2022.952922.

Bumanis, N., Kviesis, A., Paura, L., Arhipova, I., Adjutovs, M., 2023. Hen egg production forecasting: Capabilities of machine learning models in scenarios with limited data sets. Appl. Sci. 13 (13), 7607. https://doi.org/10.3390/app13137607.

Carletto, C., 2021. Better data, higher impact: Improving agricultural data systems for societal change. Eur. Rev. Agric. Econ. 48 (4), 719–740. https://doi.org/10.1093/erae/jbab030.

Coleman, T.F., Li, Y., 1996. An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim. 6, 418–445. https://doi.org/10.1137/0806023.

Ding, C., Zhao, J., Sun, S., 2023. Concept drift adaptation for time series anomaly detection via transformer. Neural Process. Lett. 55 (3), 2081–2101. https://doi.org/10.1007/s11063-022-11015-0.

Dunn, I.C., Bain, M., Edmond, A., Wilson, P.W., Joseph, N., Solomon, S., De Ketelaere, B., De Baerdemaeker, J., Schmutz, M., Preisinger, R., Waddington, D., 2005. Heritability and genetic correlation of measurements derived from acoustic resonance frequency analysis; a novel method of determining eggshell quality in domestic hens. Br. Poult. Sci. 46 (3), 280–286. https://doi.org/10.1080/00071660500098574.

Fialho, F.B., Ledur, M.G., 1997. Segmented polynomial model for estimation of egg production curves in laying hens. Br. Poult. Sci. 38 (1), 66–73. https://doi.org/10.1080/00071669708417942.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A., 2014. A survey on concept drift adaptation. ACM Comput. Surv. 46 (4), 1–37. https://doi.org/10.1145/2523813.

Gautron, J., Réhault-Godbert, S., Van de Braak, T.G.H., Dunn, I.C., 2021. What are the challenges facing the table egg industry in the next decades and what can be done to address them? Animal 15, 100282. https://doi.org/10.1016/j.animal.2021.100282.

Gonzalez-Mora, A.F., Rousseau, A.N., Larios, A.D., Godbout, S., Fournel, S., 2022. Assessing environmental control strategies in cage-free aviary housing systems: Egg production analysis and Random Forest modeling. Comput. Electron. Agric. 196, 106854. https://doi.org/10.1016/j.compag.2022.106854.

Grossman, M., Gossman, T.N., Koops, W.J., 2000. A model for persistency of egg production. Poult. Sci. 79, 1715–1724. https://doi.org/10.1093/ps/79.12.1715.

Gumiran, C.R., 2024. Fuzzy logic approach in predicting egg production on laying hen in an uncontrolled temperature. Int. J. Eng. Trends Technol. 72 (6), 146–152. https://doi.org/10.14445/22315381/IJETT-V72I6P115.

Iwana, B.K., Uchida, S., 2021. An empirical survey of data augmentation for time series classification with neural networks. PLoS One 16 (7), e0254841. https://doi.org/10.1371/journal.pone.0254841.

Ji, H., Xu, Y., Teng, G., 2025. Predicting egg production rate and egg weight of broiler breeders based on machine learning and Shapley additive explanations. Poultry Science 104 (1), 104458. https://doi.org/10.1016/j.psj.2024.104458.

Keogh, E., Ratanamahatana, C.A., 2005. Exact indexing of dynamic time warping. Knowl. Inf. Syst. 7 (3), 358–386. https://doi.org/10.1007/s10115-004-0154-9.

Kloska, M., Grmanova, G., Rozinajova, V., 2023. Expert enhanced dynamic time warping based anomaly detection. Expert Syst. Appl. 225, 120030. https://doi.org/10.1016/j.eswa.2023.120030.

Krawczyk, B., Woźniak, M., 2014. One-class classifiers with incremental learning and forgetting for data streams with concept drift. Soft. Comput. 19, 3387–3400. https://doi.org/10.1007/s00500-014-1492-5.

Li, D., Tong, Q., Shi, Z., Zheng, W., Wang, Y., Li, B., Yan, G., 2020. Effects of cold stress and ammonia concentration on productive performance and egg quality traits of laying hens. Animals (Basel) 10 (12), 2252. https://doi.org/10.3390/ani10122252.

Lokhorst, C., Lamaker, E.J.J., 1996. An expert system for monitoring the daily production process in aviary systems for laying hens. Comput. Electron. Agric. 15 (3), 215–231. https://doi.org/10.1016/0168-1699(96)00017-8.

Long, A. & Wilcox, S. 2011. Optimizing egg revenue for poultry farmers:1-10.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G., 2018. Learning under concept drift: A review. IEEE Trans. Knowl. Data Eng. 31 (12), 2346–2363. https://doi.org/10.1109/TKDE.2018.2876857.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., Fernández-Leal, Á., 2022. Human-in-the-loop machine learning: A state of the art. Artif. Intell. Rev. 56, 3005–3054. https://doi.org/10.1007/s10462-022-10246-w.

Narinc, D., Uckardes, F., Aslan, E., 2019. Egg production curve analyses in poultry science. Worlds Poult. Sci. J. 70 (4), 817–828. https://doi.org/10.1017/S0043933914000877.

Omomule, T.G., Ajayi, O.O., Orogun, A.O., 2020. Fuzzy prediction and pattern analysis of poultry egg production. Comput. Electron. Agric. 171, 105301. https://doi.org/10.1016/j.compag.2020.105301.

Ramírez-Morales, I., Cebrián, R.V., Fernandez-Blanco, E., Pazos, A., 2016. Early warning in egg production curves from commercial hens: A SVM approach. Comput. Electron. Agric. 121, 169–179. https://doi.org/10.1016/j.compag.2015.12.009.

Ramírez-Morales, I., Fernandez-Blanco, E., Rivero, D., Pazos, A., 2017. Automated early detection of drops in commercial egg production using neural networks. Br. Poult. Sci. 58 (6), 739–747. https://doi.org/10.1080/00071668.2017.1379051.

Ribeiro, R., Casanova, D., Teixeira, M., Wirth, A., Gomes, H.M., Borges, A.P., Enembreck, F., 2019. Generating action plans for poultry management using artificial neural networks. Comput. Electron. Agric. 161, 131–140. https://doi.org/10.1016/j.compag.2018.02.017.

Roberts, J.R., Souillard, R., Bertin, J., 2011. Avian diseases which affect egg production and quality. In: Improving the safety and quality of eggs and egg products. Woodhead Publishing, pp. 376–393. https://doi.org/10.1533/9780857093912.3.376.

Rodenburg, T.B., Tuyttens, F.A., Sonck, B., De Reu, K., Herman, L., Zoons, J., 2005. Welfare, health, and hygiene of laying hens housed in furnished cages and in alternative housing systems. J. Appl. Anim. Welf. Sci. 8 (3), 211–226. https://doi.org/10.1207/s15327604jaws0803_5.

Schmidl, S., Wenig, P., Papenbrock, T., 2022. Anomaly detection in time series: A comprehensive evaluation. Proceedings of the VLDB Endowment 15 (9), 1779–1797. https://doi.org/10.14778/3538598.3538602.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C., 2001. Estimating the support of a high-dimensional distribution. Neural computation 13 (7), 1443–1471. https://doi.org/10.1162/089976601750264965.

Schukat, S., Heise, H., 2021. Towards an understanding of the behavioral intentions and actual use of smart products among German farmers. Sustainability 13 (12), 6666. https://doi.org/10.3390/su13126666.

Sigognault Flochlay, A., Thomas, E., Sparagano, O., 2017. Poultry red mite (Dermanyssus gallinae) infestation: a broad impact parasitological disease that still remains a significant challenge for the egg-laying industry in Europe. Parasit. Vectors 10, 357. https://doi.org/10.1186/s13071-017-2292-4.

Statnikov, A., 2011. A gentle introduction to support vector machines in biomedicine: Theory and methods. World Scientific. https://doi.org/10.1142/792.

Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. Least squares support vector machines. World Scientific. https://doi.org/10.1142/5089.

Traore, O.Z., Doyon, M., 2023. Economic sustainability of extending lay cycle in the supply-managed Canadian egg industry. Front. Anim. Sci. 4. https://doi.org/10.3389/fanim.2023.1201771.

Tuyttens, F.A.M., Molento, C.F.M., Benaissa, S., 2022. Twelve threats of Precision Livestock Farming (PLF) for animal welfare. Front. Vet. Sci. 9, 889623. https://doi.org/10.3389/fvets.2022.889623.

Van Veen, L.A., Van den Oever, A.C.M., Kemp, B., Van den Brand, H., 2023. Perception of laying hen farmers, poultry veterinarians, and poultry experts regarding sensor-based continuous monitoring of laying hen health and welfare. Poult. Sci. 102 (5), 102581. https://doi.org/10.1016/j.psj.2023.102581.

Woudenberg, S.P.D., Van der Gaag, L.C., Feelders, A., Elbers, A.R.W., 2014. Real-time adaptive problem detection in poultry. Front. Artif. Intell. Appl. 263, 1217–1218. https://doi.org/10.3233/978-1-61499-419-0-1217.

Yamak, U.S., Sarica, M., 2012. Relationships between feather score and egg production and feed consumption of different layer hybrids kept in conventional cages. Archiv fur Geflügelkunde 76, 31–37.

Yin, H., Wu, Z., Wu, J.-C., Chen, Y., Chen, M., Luo, S., Gao, L., Hassan, S.G., 2023. A multistep interval prediction method combining environmental variables and attention mechanism for egg production rate. Agriculture 13 (6), 1255. https://doi.org/10.3390/agriculture13061255.

Yokkampon, U., Chumkamon, S., Mowshowitz, A., Fujisawa, R., Hayashi, E., 2021. Anomaly detection using support vector machines for time series data. J. Rob. Netw. Artif. Life 8 (1), 41–46. https://doi.org/10.2991/jrnal.k.210521.010.