

Article

A Machine Learning Framework Based on Extreme Gradient Boosting to Predict the Occurrence and Development of Infectious Diseases in Laying Hen Farms, Taking H9N2 as an Example

Yu Liu ^{1,2}, Yanrong Zhuang ³, Ligen Yu ^{1,2,*}, Qifeng Li ^{1,2}, Chunjiang Zhao ^{1,2}, Rui Meng ^{1,2}, Jun Zhu ^{1,2} and Xiaoli Guo ^{1,2}

¹ Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

² National Innovation Center of Digital Technology in Animal Husbandry, Beijing 100097, China

³ College of Water Resources and Civil Engineering, China Agricultural University, Beijing 100083, China

* Correspondence: yulg@nercita.org.cn; Tel.: +86-10-51503855

Simple Summary: The H9N2 avian influenza virus has spread to the whole world and become one of the dominant subtype influenza viruses in chickens in China. H9N2 virus could not only result in great economic losses by reducing egg production but also serve as a gene vector to provide its gene segments to other emerging severe influenza A viruses to cause higher mortality and more serious consequences. Thus, developing models to predict H9N2 status should be given priority. Our main aim was to use the machine learning method (XGBoost classification algorithm) and regular production data (laying rate and mortality) to predict the occurrence and development of H9N2 in laying hen houses and evaluate their performance in disease prediction. Additionally, we assessed the working ability of the framework with different time frames to predict H9N2 status in advance within a 3-day period. We found that this framework could work well in prediction with high accuracy and sensitivity, and with more information introduced into the model, more “don’t care values” would be added into datasets to affect model performance by forming attribute noise. Besides, our study recommended efficient frameworks and models for H9N2 status prediction and provided practical potential uses in the livestock and poultry industry.

Abstract: The H9N2 avian influenza virus has become one of the dominant subtypes of avian influenza virus in poultry and has been significantly harmful to chickens in China, with great economic losses in terms of reduced egg production or high mortality by co-infection with other pathogens. A prediction of H9N2 status based on easily available production data with high accuracy would be important and essential to prevent and control H9N2 outbreaks in advance. This study developed a machine learning framework based on the XGBoost classification algorithm using 3 months’ laying rates and mortalities collected from three H9N2-infected laying hen houses with complete onset cycles. A framework was developed to automatically predict the H9N2 status of individual house for future 3 days (H9N2 status + 0, H9N2 status + 1, H9N2 status + 2) with five time frames (day + 0, day – 1, day – 2, day – 3, day – 4). It had been proven that a high accuracy rate > 90%, a recall rate > 90%, a precision rate of >80%, and an area under the curve of the receiver operator characteristic ≥ 0.85 could be achieved with the prediction models. Models with day + 0 and day – 1 were highly recommended to predict H9N2 status + 0 and H9N2 status + 1 for the direct or auxiliary monitoring of its occurrence and development. Such a framework could provide new insights into predicting H9N2 outbreaks, and other practical potential applications to assist in disease monitor were also considerable.

Keywords: H9N2 status; laying hen house; machine learning; predictive performance



Citation: Liu, Y.; Zhuang, Y.; Yu, L.; Li, Q.; Zhao, C.; Meng, R.; Zhu, J.; Guo, X. A Machine Learning Framework Based on Extreme Gradient Boosting to Predict the Occurrence and Development of Infectious Diseases in Laying Hen Farms, Taking H9N2 as an Example. *Animals* **2023**, *13*, 1494. <https://doi.org/10.3390/ani13091494>

Academic Editor: Esmail Ebrahimic

Received: 4 March 2023

Revised: 26 April 2023

Accepted: 26 April 2023

Published: 27 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, various infectious diseases have posed risks to the livestock and poultry industry [1], such as avian influenza, foot and mouth disease (FMD), and classical swine fever. Infectious diseases are mostly caused by viruses, bacteria, fungi, etc. [2], which could rapidly outbreak and spread to greatly limit and test the epidemic prevention and response capacity of government and farmers. To date, infectious disease outbreaks have caused tremendous socioeconomic losses and public health consequences globally due to disease control measures [3], such as culling animals in endemic areas, temporary bans on the import and export of livestock products, and the limitation of livestock production and market activities [4].

For infectious disease management, the most appropriate strategy is accurate and rapid detection and control [5]. Researchers have developed some novel approaches of classifications based on association models to identify virus sequences, virus host range and zoonotic transmissible sequences to avoid pandemic or disastrous epidemics [6,7]. Infectious disease prediction has great value for initiating rapid responses and for the safety of the livestock and poultry industry. In order to well monitor animal health and diseases, many prediction models for infectious diseases have been developed in recent years [8] for early warning of individual animals, animal herds, and regions. Prediction models based on classical statistical and mathematical methods (e.g., logistic regression) have been widely demonstrated in epidemiological risk warning [9], which were usually considered as kind suitable techniques for identifying risk factors (e.g., lameness recognition in broiler [10]) rather than establishing outbreak prediction models [9,11]. There is increasing interest in trying to use machine learning methods for epidemic prediction [11,12], especially the tree-based machine learning methods, such as random forest (RF) and extreme gradient boosting (XGBoost), which have been widely used in predicting diseases with the advantages of high efficiency, and strong interpretability [13–15]. In the field of veterinary sciences, only a few studies focused on the prediction of disease outbreaks. RF was demonstrated to have good capability in predicting FMD outbreaks with the external risk factors as input, like purchasing, neighbors, vaccinations and close to the main road [12], and a knowledge map was maturely applied to diagnose livestock and poultry diseases with varied symptoms [16,17]. Obviously, additional time and labor are required to collect and count these input variables mentioned above in addition to the normal production data. It would be more convenient and beneficial for farmers and veterinarians to identify animal health abnormalities and rapidly respond if the regular production data could be used directly to reflect disease outbreaks, but it is still a gap. Meanwhile, the developed models just predicted the current outbreak status and were not forward-looking. Many studies have been using a time-series forecasting model to explain, evaluate and estimate the development trends in outbreak human diseases (e.g., COVID-19), such as the further values of outbreak disease cases, deaths, or transmission rates [18,19]. Besides, XGBoost was proved to have more outstanding prediction accuracy and generalization ability than RF, support vector machines (SVM), k-nearest neighbor (KNN), and back propagation neural network (BPNN) in predicting the risk and outbreaks of diabetes [20], dermatomyositis [21], COVID-19 [14,15], wheat stripe rust [22], and other diseases of human and plants, due to the superiority of its architecture, which also provided a new idea and algorithm for risk/outbreak identification, early intervention, and practical application in animal diseases.

Different infectious diseases do not only show different symptoms in pathological changes but also show different effects on production performance indicators due to differences in virulence, adaptability and replication; even different strains of the same disease can have different clinical manifestations [23]. For example, there were significant differences in the survival percentage and weight-changing trends of broilers infected with HN10PY01 and NX0101 of avian leukosis virus-J field strains [23]. One study showed the survival rate was 30%, 50%, and 90% in broilers inoculated with *Ornithobacterium rhinotracheale* (ORT) + H9N2 avian influenza virus, ORT and H9N2 virus alone, respec-

tively [24]. These indicators mentioned above are the regular ones recorded every day in most commercial farms, which supports the use of regular production data to reflect different diseases. Even though some different diseases show similarities in production data, the prediction of infectious disease outbreaks with regular production data will also be beneficial to help governments and farmers rapidly respond to screening and control infectious diseases.

Generally, effective detection and monitoring are necessary for infectious disease management [25]. This study used the data from laying hens infected with the H9N2 virus as an example, trying to use a machine learning method (XGBoost classification algorithm) and regular production data (laying rate and mortality) to predict the occurrence and development of H9N2 and prove the prediction performance of the framework and XGBoost models to provide an efficient method for infectious diseases early warning in livestock and poultry farms.

2. Materials and Methods

Avian Influenza caused by H9N2 avian influenza virus (“H9N2 Avian Influenza” referred to as “H9N2” below) can pose a serious threat to both the poultry industry and public health safety, and an effective and rapid diagnostic method for H9N2 occurrence and development is urgently needed for prompt prevention and control.

2.1. H9N2 Cases and Data Collection

This study collected information on H9N2 occurrence and development and correlated production data from a large-scale concentrated laying hen farm located in Beijing. Trained and experienced veterinarians, who routinely conduct disease identification and abnormality diagnosing, helped to confirm the H9N2 cases (including the start and end of this epidemic) with their experiences and the results of the avian influenza virus antigen detection card based on the colloidal gold method.

Data were collected from 3 infected laying hen houses every day from the period of October to December 2012, covering the whole onset cycle of each outbreak. The numbers of days of having H9N2 and non-H9N2 were 37 and 88, with a case-to-control ratio of approximately 1:2, fitting the typically recommended ratio of observational studies of disease incidence [26]. Variables collected in this study are listed in Table 1. Temperature and relative humidity were automatically measured by sensors and recorded in the environment management system. Date, age, laying rate and mortality were automatically generated, calculated and recorded by the management system. The number of abnormal laying hens, the qualified rate of immune antibodies against highly pathogenic avian influenza (HPAI) (QR_{immu}), and the treatment for H9N2 infection were confirmed, recorded, and provided by a veterinarian. Veterinarians treated the laying hens with drugs from the day that the first H9N2-infected bird was detected until a non-H9N2-infected bird was confirmed.

Table 1. Variables about the information of H9N2 occurrence and development and correlated production data.

Variables	Variables
Date	Laying rate/%
Age/days	Mortality/%
Maximum temperature throughout the day (T_{max})/°C	Number of abnormal laying hens ($N_{\text{ab-hens}}$)
Minimum temperature throughout the day (T_{min})/°C	Qualified rate of immunization (QR_{immu})/%
Average relative humidity (RH)/%	Treatment for H9N2 infection

2.2. Xgboost Algorithm

XGBoost is 1 of the latest tree-boosting algorithms based on a novel sparsity-aware algorithm and weighted quantile sketch to develop a “strong” learner using all prediction

results of “weak” learners [27,28] with great success in terms of both performance and speed. As a tree-based model, the XGBoost classification algorithm changes the score of leaves to a class tag and finally outputs the class tag with the highest proportion of allocated leaves. In this study, the XGBoost classification model was built to predict H9N2 status to reflect the occurrence and development process of this epidemic, and Figure 1 uses an example to illustrate the progress of the XGBoost classification model using 2 trees, each with 2 depths to predict the status. In fact, there are always more than 2 trees and 2 depths in model training. H9N2 status was dealt with binary problems with the labels of 0 and 1, of which “0” meant “no H9N2 case”, and “1” meant “having H9N2 case”. If the sum of the function leaf scores was greater than 0.5, it was judged to belong to 1; otherwise, it belonged to 0. Features including laying rate, mortality and time frame features, which were directly and closely related to H9N2 outcomes, were used as the input.

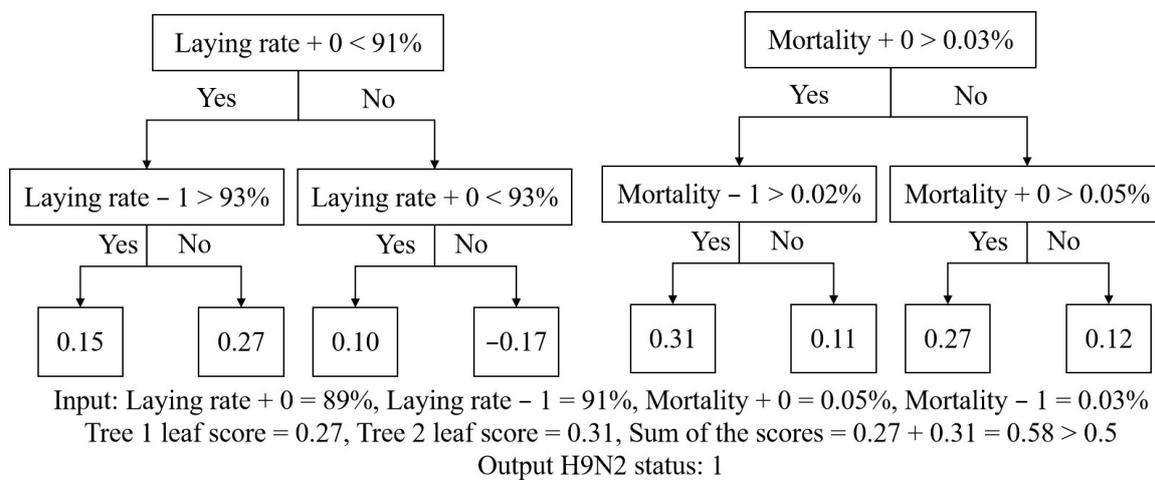


Figure 1. Illustration of XGBoost classification model using 2 trees, each with 2 depths to predict H9N2 status, an example. Laying rate + 0 = laying rate of today, laying rate - 1 = laying rate of yesterday, mortality + 0 = mortality of today, mortality - 1 = mortality of yesterday.

Hyperparameters, such as the regularization parameters in XGBoost, can greatly affect the performance of a model [29]. In this study, Grid Search Cross-Validation exhaustive parameter search technique with fivefold cross-validation was applied to search the optimal hyperparameters involved in the performance of the models, which included *learning_rate*, *n_estimators*, *max_depth*, *min_child_weight*, *gamma*, *subsample*, and *colsample_bytree*, with the search spaces of 0.001~0.2, 0~100, 1~15, 1~15, 0~1, 0.7~0.9, and 0.7~0.9, respectively. The optimized hyperparameters were determined by the maximum value of the accuracy. Then, the importance of these features in building and boosting the tree-based XGBoost classification model was calculated for model interpretation [30].

2.3. Framework Design

All of the XGBoost classification algorithm running, hyperparameters tuning, and statistical analysis were conducted using Python software 3.6 and the scikit-learn toolkit. The examples of Python codes and collated datasets were provided as supplemental files (Codes S1 and Datasets S1).

The overall workflow for the prediction task is given in Figure 2. In order to prove the model’s good application ability, the collected datasets from 2 infected laying hen houses were used as a training set (about 64% of all data), and datasets from the 3rd infected laying hen house were used as a testing set (about 36% of all data). Fivefold cross-validation was conducted in each model training process to evaluate the performance of such a framework under practical farm management conditions. The hyperparameters were tuned only utilizing the training set. The input factors included the laying rate with time frames and the mortality with time frames, and the output value was H9N2 status. After training

XGBoost models, the accuracy, recall rate, precision rate and receiver operator characteristic (ROC) curve of each model were calculated to evaluate the framework performance.

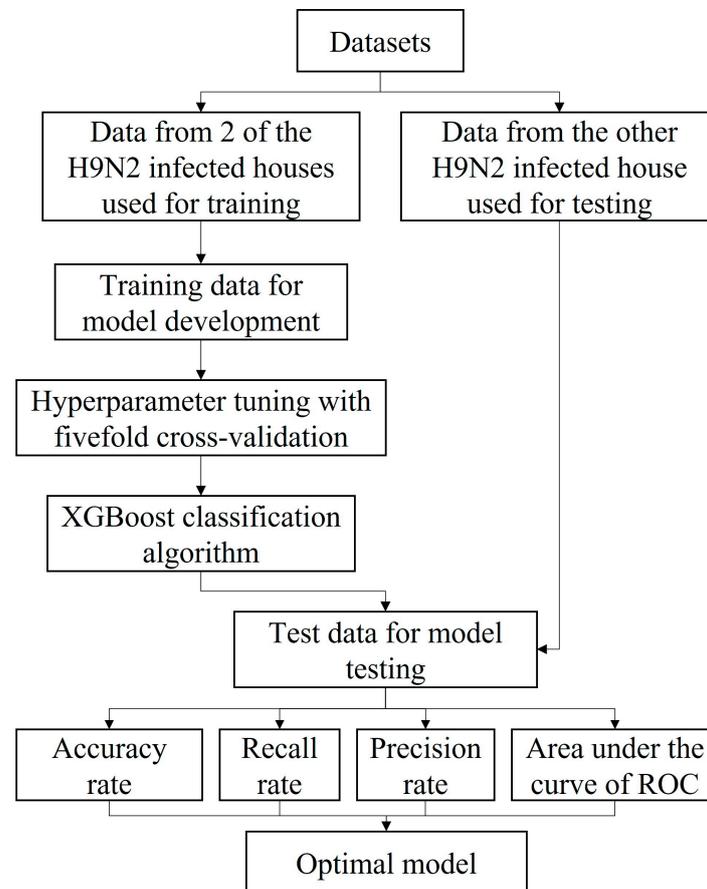


Figure 2. Overall workflow of model development for H9N2 status prediction. ROC, receiver operator characteristic curve.

Laying hens inevitably suffer different diseases or breeding equipment failures in the normal breeding process, which will affect the normal laying rate and mortality changing regularly. In order to avoid the impact of outliers in long-term production data on the classification performance of the model and to reduce the data requirement for small sample events such as infectious disease outbreaks, this study tried to investigate the suitable time frame using for H9N2 status confirming. Given that H9N2 can cause sudden changes in production performance, choosing long-term data is not obviously helpful for the prediction of disease outbreaks; this study set 5 time frames rolling windows of today (day + 0), the previous 1 day (day − 1), the previous 2 days (day − 2), the previous 3 days (day − 3), and the previous 4 days (day − 4) to study the suitable time frame for H9N2 status confirming. In each time frame case, 3 XGBoost models were built to predict coupled future 3 days' values, of which model #1 would predict today's status (i.e., H9N2 status + 0), model #2 would predict the next day's status (i.e., H9N2 status + 1), and model #3 would predict the future 2nd day's status (i.e., H9N2 status + 2). The H9N2 status was defined into 2 categories, "no" (marked as 0) and "yes" (marked as 1), of which "no" meant no H9N2 case, and "yes" meant having an H9N2 case. Due to each prediction being targeted on a 3-day period, a time series 3-day rolling window was applied to create input and output variables in train or test datasets, as shown in Figure 3.

Time frame	1 October 2012	2 October 2012	3 October 2012	4 October 2012	5 October 2012	6 October 2012	7 October 2012	8 October 2012
Day + 0	Input	Output	Output					
Day - 1	Input	Input	Output	Output	Output			
...Different time frames...								
Day - 4	Input	Input	Input	Input	Output	Output	Output	Output

Input
 Output

Figure 3. Examples of different time frames to simulate the execution of the XGBoost classification model framework. +0 = today, −1 = previous 1 day, −2 = previous 2 days, −3 = previous 3 days, and −4 = previous 4 days.

2.4. Evaluation Criteria

There were 4 generally used criteria calculated to evaluate and compare the prediction capacity of developed models with different time frames. This study defined the 0 status (no H9N2 case) as negative and the 1 status (having H9N2 case) as positive. The agreement of the results between the actual and predicted status was described by the number of true positives (*TP*), true negatives (*TN*), false positives (*FP*), and false negatives (*FN*), of which the sum of *TP*, *TN*, *FP* and *FN* was the total number of observed days in 3rd laying hen house (*N*). Accuracy rate, recall rate, precision rate, and the area under the curve (AUC) of the ROC were used to assess the performances of the XGBoost models [12]:

- (1) Accuracy rate (ACC)

$$\text{ACC} = \frac{TP + TN}{N} \times 100\% \quad (1)$$

- (2) Recall rate (RR)

$$\text{RR} = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

- (3) Precision rate (PR)

$$\text{PR} = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

AUC of ROC

AUC is an indicator of the discriminative or predictive ability of prediction models. The prediction ability could be generally judged based on the AUC values: AUC = 0.5, no discrimination; $0.5 < \text{AUC} < 0.6$, poor discrimination; $0.6 \leq \text{AUC} < 0.7$, fair discrimination; $0.7 \leq \text{AUC} < 0.8$, acceptable discrimination; $0.8 \leq \text{AUC} < 0.9$, excellent discrimination; and $\text{AUC} \geq 0.9$, outstanding discrimination [31]. A higher AUC value means the model prediction performance is better.

3. Results

3.1. Descriptive Statistic

The descriptive statistics associated with the variables about the information of H9N2 occurrence and development and correlated production data are shown in Table 2. The age range of studied birds was 161–266 days, with an average laying rate of 93.3% and an average mortality of 0.03% during the normal production period. Their average age was around 27 weeks, which was in the peak laying period [32]. During the period of data collection, the temperature ranges of houses 1 and 2 were similar, with T_{\max} values from

3 to 24 °C and T_{min} values from -10 to 9 °C; the temperature of house 3 was significantly lower than the other 2 houses ($p < 0.05$). The average RH range of 3 houses was from 35 to 65%. With the QR_{immu} above 80%, the extreme values of laying rate, mortality, and $N_{ab-hens}$ were 84.1%, 0.20%, and 265 hens, respectively, after the hens were infected with H9N2.

Table 2. Variable description of the overall data in the 3 H9N2 infected laying hen houses.

	Variable	Unit	Max	Mean	Min	SD
House 1	Age	days	199	-	161	-
	T_{max}	°C	24	15	5	5
	T_{min}	°C	9	1	-6	4
	RH	%	65	52	38	7
	Laying rate	%	97.1	92.5	87.4	2.9
	Mortality	%	0.10	0.03	0.01	0.02
	$N_{ab-hens}$	hens	201	-	0	-
	QR_{immu}	%	90	90	90	0
House 2	Age	days	258	-	218	-
	T_{max}	°C	22	11	3	5
	T_{min}	°C	9	-2	-10	4
	RH	%	65	49	35	8
	Laying rate	%	94.3	92.4	86.1	2.3
	Mortality	%	0.20	0.04	0.01	0.03
	$N_{ab-hens}$	hens	265	-	0	-
	QR_{immu}	%	80	80	80	0
House 3	Age	days	266	-	222	-
	T_{max}	°C	12	4	-4	5
	T_{min}	°C	1	-6	-14	4
	RH	%	65	46	35	7
	Laying rate	%	93.4	91.4	84.1	2.7
	Mortality	%	0.13	0.05	0.01	0.03
	$N_{ab-hens}$	hens	214	-	0	-
	QR_{immu}	%	80	80	80	0

The data from house 1 was taken as an example to reflect the basic changing of environment and production variables in Figure 4. It was clear that the temperature showed a slow downward trend; RH fluctuated regularly and was stable in total. Before the H9N2 outbreak, the laying rate increased gradually and stabilized at 96% according to the normal growth curve of laying hens. Meanwhile, the mortality regularly changed to around 0.02%. After the H9N2 outbreak, both the laying rate and mortality were sharply affected. Interestingly, there were slight abnormal changes in laying rate and mortality occurring 1 or 2 days before the outbreak, which could be mined to be the portends of the H9N2 outbreak.

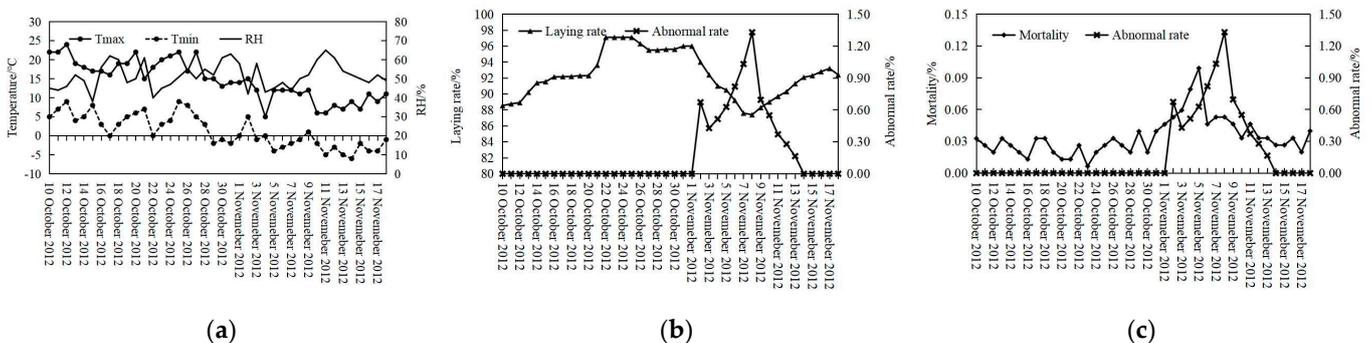


Figure 4. Taking house 1 as an example to show the changes in environment and production variables during the data collection period: (a) T_{max} , T_{min} , and RH; (b) Laying rate and abnormal rate; (c) Mortality and abnormal rate.

3.2. Predictive Performance Comparison

Table 3 presents the comparative results of the model performances with different time frames to predict 3 future days' H9N2 status. All the prediction models had accuracy rates greater than 80% with high recall rates, which meant the high sensitivity of these models. It could be seen day + 0 and day – 1 classification models had the same accuracy and sensitivity in predicting H9N2 status + 0, +1 and +2. With the accumulated days used as input data increasing, all of the ACC, RR and PR fluctuated and decreased. The XGBoost classification models predicted the likelihood of the H9N2 status + 0 most precisely since they revealed the highest PRs. ACC and RR of models for H9N2 status + 0 kept stable at 92.31% with the accumulated days from today to the previous 3 days, then they decreased with the input days increasing. It was worth noting that the performance of the day – 2 model was the lowest among all of the models to predict H9N2 status + 1; other models had the same ACC and RR. Meanwhile, for H9N2 status + 2 prediction, all of the models had the same performances of ACC and RR. It was clear that when the models predicted the H9N2 status of more future days, the performances were worse in general. The performances of the models of days + 0, –1, –2, and –3 to predict H9N2 status + 0 were the best, with an ACC and RR of 92.31%, and H9N2 status + 1 was the next.

Table 3. Model performance parameters, including accuracy (ACC), recall rate (RR) and precision rate (PR) from 15 XGBoost models with 5 time frames of today (day + 0), previous 1 day (day – 1), previous 2 days (day – 2), previous 3 days (day – 3), and previous 4 days (day – 4) to predict future 3 days' H9N2 status (+0, +1, +2).

Criteria	Predicted Future Days	Day + 0	Day – 1	Day – 2	Day – 3	Day – 4
ACC/%	H9N2 status + 0	92.31	92.31	92.31	92.31	89.74
	H9N2 status + 1	89.74	89.74	87.18	89.74	89.74
	H9N2 status + 2	84.62	84.62	84.62	84.62	84.62
RR/%	H9N2 status + 0	92.31	92.31	92.31	92.31	84.62
	H9N2 status + 1	92.31	92.31	84.62	92.31	92.31
	H9N2 status + 2	84.62	84.62	84.62	84.62	84.62
PR/%	H9N2 status + 0	85.71	85.71	85.71	85.71	84.62
	H9N2 status + 1	80.00	80.00	78.57	80.00	80.00
	H9N2 status + 2	73.33	73.33	73.33	73.33	73.33

A higher AUC value shows a better prediction performance of the evaluated model. Analyzing Figure 5, the AUC values of most models were higher than 0.90, indicating their outstanding discrimination. Regarding the AUC values, similar to the results from Table 3, the more days the models predicted, the worse the performances were. The performance of models predicting H9N2 status + 0 was the best, with all the AUC values higher than 0.90, followed by H9N2 status + 1 and H9N2 status + 2. When the input data came from more days, the AUC values would become worse. ACC and RR of models for H9N2 status + 0 kept stable at 92.31% with the accumulated days from today to the previous 3 days, then they decreased with the input days increasing. Models of day + 0 and day – 1 to predict the H9N2 status of different future days had similar great performances, and the highest AUC value was found in the day – 1 model for H9N2 status + 0 with an AUC of 0.98.

3.3. Feature Importance

Due to the best performances of models to predict H9N2 status + 0, Figure 6 takes H9N2 status + 0 prediction models as examples to analyze the feature importance ranking, which refers to the contributions of individual input variables to the performance of the model. In most of these models, laying rate and mortality were identified as playing different roles in building models, except the day + 2 model for H9N2 status prediction, where only laying rates were selected in building the model. This indicated the variations

in laying production and mortality between different days were the key driver in training models. The results revealed that laying rate was the more important feature in predicting H9N2 status than mortality. If the production data only from today and the previous day were used as input, the laying rate + 0 showed its great importance in prediction models. While, with the date from more accumulated days inputting, laying rate + 0 seemed not as important as before, the laying rates of the previous 1 and more days played the key role in prediction. Similarly to mortality, the importance of mortality from today and the adjacent days was lower and lower. For instance, the mortality from different days had similar importance and was ranked last.

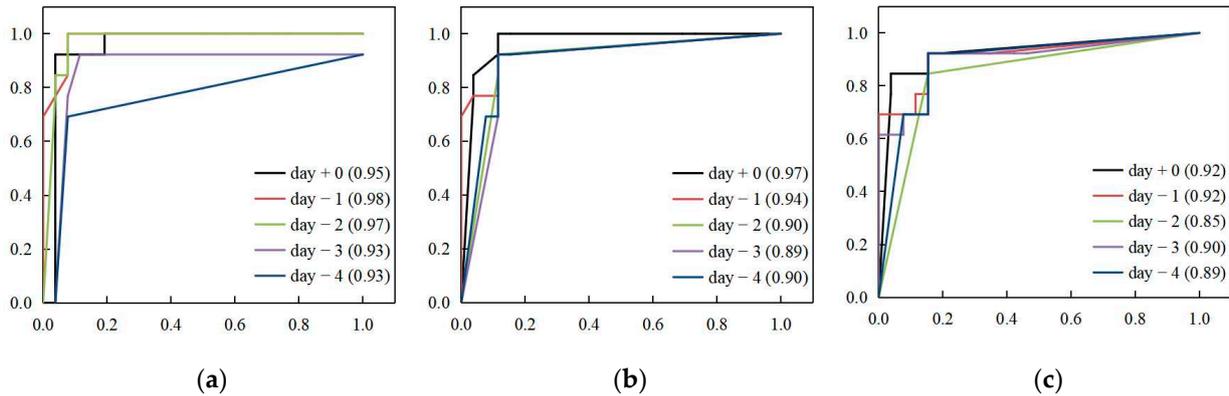


Figure 5. Receiver operator characteristic (ROC) curve from XGBoost models to predict H9N2 status of today (a), next day (b), and the future 2nd day (c) in laying hen houses with time frames of day + 0, day - 1, day - 2, day - 3, and day - 4. The value enclosed in parentheses meant the area under the curve. +0 = today, -1 = previous 1 day, -2 = previous 2 days, -3 = previous 3 days, and -4 = previous 4 days.

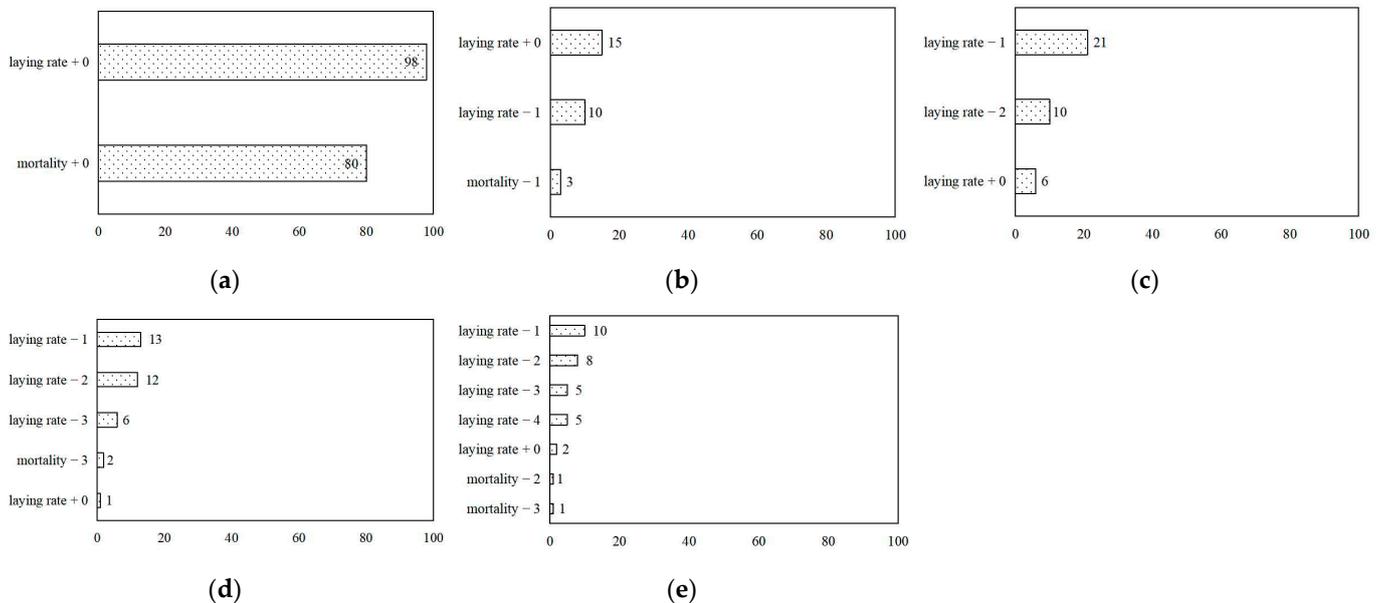


Figure 6. Feature importance for H9N2 status + 0 prediction in laying hen houses from the XGBoost classification models with time frames of day + 0 (a), day - 1 (b), day - 2 (c), day - 3 (d), and day - 4 (e). +0 = today, -1 = previous 1 day, -2 = previous 2 days, -3 = previous 3 days, and -4 = previous 4 days.

4. Discussion

In the present study, the XGBoost classification algorithm was applied to H9N2-related data to develop models designed to predict the H9N2 status among laying hen houses using temporal variables and production variables as predictors. The predictive performances of XGBoost models were evaluated and compared.

4.1. Relationship between Environment and Production Variables and H9N2 Status

Studies proved low temperatures would help avian influenza viruses survive in the environment to facilitate viruses' introduction and outbreaks [33]. Meanwhile, the low temperature would decrease the hens' immunity, which could potentially cause the H9N2 outbreak in the laying hen house. As shown in Table 2 and Figure 4, the temperature gradually dropped even to $-14\text{ }^{\circ}\text{C}$, which was far away from the optimum temperature (thermoneutral zone, $19\text{--}22\text{ }^{\circ}\text{C}$) for laying hens [34]. And it could be proved by the correlation analysis in Table 4.

Table 4. Correlation analysis of environment and production variables and H9N2 status (yes or no) in 3 laying houses.

Laying Hen House	T _{max}	T _{min}	RH	Laying Rate	Mortality
House 1	−0.61 **	−0.45 **	−0.01	−0.58 **	0.67 **
House 2	−0.30	−0.36 *	−0.31 *	−0.86 **	0.66 **
House 3	−0.67 **	−0.52 **	−0.21	−0.87 **	0.83 **

* Significant difference at 0.05 level. ** Significant difference at 0.01 level.

Since the hens were infected, H9N2 would spread to the whole house, and H9N2 had risen to be the most prevalent subtype of avian influenza virus in China, causing great economic losses because of the reduced egg production [35,36]. Clearly, after the H9N2 outbreak, the laying rate sharply decreased and was significantly and negatively correlated with H9N2 positive status ($p < 0.01$). Besides, H9N2 would cause high mortality associated with co-infection with other pathogens [37], such as H5N2 [38], H6N1 [39], H7N7 [38], H7N9 [40], and H10N8 [41]. H9N2 could also cause mortality due to the pathogenicity itself or the enhancement of other disease responses. In Korea, when layers were infected with H9N2, the mortality was about 30%, with reduced egg production [42]. In our study, all of the laying hens were vaccinated with immune antibodies against HPAI, with a qualification rate of above 80%. This H9N2 epidemic just caused a small increase in mortality but still significantly and positively affected the mortality ($p < 0.01$).

Furthermore, 1 or 2 days before the H9N2 had been detected, the laying rate and mortality changed abnormally compared with the data from the days before. In this stage, H9N2 was not detected, which might be because of its low concentration, or because it caused such little impact on production to fail to raise the alarm among farmers and veterinarians. For epidemics, a delay of 1 or 2 days can lead to a wider impact on animal health and performance, especially the H9N2 can serve as vehicles by donating their gene segments to other influenza A viruses [37] to cause more serious consequences. Thus, developing prediction models to control the circulation of H9N2 should be paid more attention to, and the laying rate and mortality as the priority corresponding factors could be used as input variables for H9N2 status prediction models.

4.2. Framework Performance and Model Interpretation

Several studies have reported computational fluid dynamics modeling, air dispersion modeling, and transmission network modeling on HPAI using past outbreaks of infectious viral animal diseases to determine the contribution of airborne transmission towards HPAI spread [43] during its outbreaks. As one of the important subtypes of the avian influenza virus, H9N2 could donate gene segments to other HPAI viruses. Thus, developing models to predict and control the circulation of H9N2 should be given priority.

Different from analytical models used in finding the relationship between different variables [44], the epidemic-based predictive model is focused on reliable and accurate prediction under real production. This study evaluated the XGBoost classification models with different time frames for predicting the H9N2 status of several future days. This setting ensured the model's development and future prediction were only dependent on historical information. By analyzing the performance criteria, the trend and fluctuation of ACC, RR, PR, and AUC can be seen in Table 3 and Figure 5, and it was found that the accuracy, sensitivity, and AUC values of the models to predict H9N2 status + 0 were higher than those to predict H9N2 status + 1 and +2, the drop of accuracy was 7.69% with the prediction moving forward from 0th to 2nd day. The overall accuracy rate and recall rate of this framework to predict H9N2 status + 0 with the time frame of day + 0, day - 1, day - 2 and day - 3 were demonstrated as reliable toward 92% of the predictions, and the framework to predict H9N2 status + 1 with the time frame of day + 0 and day - 1 still worked well. Besides, the data captured herein represent the before/after H9N2 occurrence and development in concentrated and HPAI-immunized laying hen houses; in order to prove the good applications of our results, data from two of the infected houses were used in model training, and data of the third house was used in model testing. Furthermore, it would be beneficial to apply the models to larger datasets collected from other endemic areas in recent years.

Once the animal suffers from a disease, the production performance usually changes suddenly. Unlike time series cross-validation models, this study chose five time frames with a fixed number of days rather than the accumulated days to predict H9N2 status. Compared with ACC, RR, PR, and AUC of different models, the models of time frames with more days generally showed worse and fluctuant performances in predicting H9N2 status + 0 and +1, which might be due to more additional and useless information introduced into the model training. For the prediction of a binary outcome, imbalanced classification often occurs when many outputs belong to one of two classes [45]. As shown in Figure 6, the laying rate and mortality of day - 3 and day - 4 played small roles in each model, and some features were even negligible. This finding was supported by the study indicating that the more information introduced into the model, the more "don't care values" would be added into datasets, which was easy to form attribute noise affect model performance [46]. The feature with a higher value indicates its greater power to explain the variation of the output and its key role in training the model [47]. Laying rate - 1 and laying rate - 2 ranking the first two were the top two important predictors in models with day - 2, -3, -4 instead of laying rate + 0 as the key predictor in models with day + 0 and day - 1, illustrating laying rate + 0 had the significant and stronger explanation for epidemic status. Similarly, for H9N2 status + 2, the ACC and RR of all models were 84.62%, and their PRs were 73.33% because their first 2 important features were laying rate + 0 and laying rate - 1 to show their similar explanation ability except day + 0 model, which only had 2 features ranking by laying rate + 0 and mortality + 0 (Figure S1).

Furthermore, in XGBoost classification models, the laying rate had higher importance than mortality features; this was consistent with the conclusions of the impact of H9N2 on poultry production [36], where the H9N2 mainly resulted in directly reducing egg production or high indirect mortality associated with the co-infection with other pathogens. In this study, XGBoost classification models of day + 0 and day - 1 to predict H9N2 status + 0 and H9N2 status + 1 were recommended for H9N2 occurrence and development predicting. Besides, there were a lot of machine learning methods used in animal disease predictions [12,48], and Gradient-Boosted Tree (GBT) was proven to be the most accurate model (accuracy of 84.9%) than Deep Learning, Decision Tree, and RF in predicting sub-clinical bovine mastitis [48]. XGBoost, as a quick implementation of GBT [49], was proved to have more outstanding prediction accuracy and generalization ability than RF, support vector machines (SVM), k-nearest neighbor (KNN), and back propagation neural network (BPNN) in predicting the risk and outbreaks of human diseases [14,15,20,21], due to its parallel processing [30], handle and capture missing values internally [50]. The

results of our study also showed good application and performance in animal infectious disease predictions.

4.3. Potential Applications

This study provided an acceptable and reliable framework to predict H9N2 status from +0th to +2nd days with a high accuracy rate, recall rate and AUC within the cross-validation. Different from the prediction generated by traditional production curves [51,52], this framework was more dynamic and individual-house-oriented, which could support many innovative applications. Basically, authorities or veterinarians may use XGBoost classification models to communicate the risk of H9N2 occurrence and development to farmers and stakeholders, and the warning rule could be described as Table 5 or revised based on the veterinary's criteria. The status of two consecutive days can be used to reflect the occurrence and development of H9N2 and give the corresponding warning level to remind and assist authorities or veterinarians in diagnosing and controlling the disease. Therefore, precise treatment can be given, such as medicating, culling or temporary bans [4]. The same framework could also be applied to early warning and auxiliary diagnosis of similar diseases that have impacts on production performance or other production traits responded in advance, such as infectious laryngotracheitis [53] and fowlpox [54]. Alternatively, the applications of these XGBoost classification models can be integrated with computer-based animal epidemic surveillance systems (e.g., the web-based FMD prediction system of Thailand [12]) or farm management systems as an additional application for H9N2 status prediction. The likelihood of an H9N2 outbreak for a particular laying hen house can be determined through these models whenever the new real-time data are recorded in the management system based on the judgment rule listed in Table 5 and through the color of the indicator on the screen to intuitively reflect H9N2 status in laying hen house.

Table 5. Warning rule of H9N2 status using XGBoost classification models.

Predicted H9N2 Status		Warning Level	Meaning	H9N2 Development
H9N2 Status + 0	H9N2 Status + 1			
0	0	Green	Safe	None
0	1	Yellow	Low warning	Start
1	1	Red	High warning	Development
1	0	Yellow	Low warning	Nearly finish

Note: +0 = today, +1 = future 1st day.

At present, Python software used in this study is free and open access, which has the ability to handle different types of data, from small to large datasets. Moreover, many open-source machine learning platforms are available and easy to operate only if the users have basic computer or statistical programming knowledge [11].

5. Conclusions

This study developed a framework using XGBoost classification models and demonstrated that their abilities with different time frames were capable of predicting H9N2 status in individual laying hen houses in advance within a 3-day period. H9N2 status prediction models worked well with accuracy and sensitivity ranging from excellent to outstanding, of which an accuracy rate > 90%, a recall rate > 90%, a precision rate > 80%, and an AUC \geq 0.85 could be achieved. Among them, the models of day + 0 and day - 1 were highly recommended to predict H9N2 status + 0 and H9N2 status + 1 for direct or auxiliary monitoring of the occurrence and development of H9N2. Authorities and veterinarians may consider applying the XGBoost classification algorithm to the currently used computer-based animal epidemic surveillance system or farm management system to enhance its early warning capability. Furthermore, the framework utilized in this study

may also be considered for application to data from other infectious diseases, which have similar impacts on production performances.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ani13091494/s1>, Codes S1: Python codes. Datasets S1: Part of the collated datasets used in this study. Figure S1: Feature importance for H9N2 status + 0 (A), +1 (B), +2 (C) predictions in laying hen houses from the XGBoost classification models with time frames of day + 0 (a), day − 1 (b), day − 2 (c), day − 3 (d), and day − 4 (e). +2 = future 2nd day, +1 = future 1st day, +0 = today, −1 = previous 1 day, −2 = previous 2 days, −3 = previous 3 days, and −4 = previous 4 days.

Author Contributions: Conceptualization, Y.L., L.Y., Q.L. and C.Z.; Methodology, Y.L. and Y.Z.; Validation, Y.Z.; Formal analysis, Y.L.; Investigation, Y.L., R.M. and J.Z.; Resources, L.Y.; Data curation, Y.L.; Writing—original draft, Y.L.; Writing—review and editing, L.Y., Q.L. and C.Z.; Visualization, Y.L., Y.Z. and X.G.; Supervision, L.Y.; Funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific and Technological Innovation 2030 Program of China Ministry of Science and Technology (2021ZD0113803), Beijing Innovation Consortium of Digital Agriculture Research System (BAIC10-2022, BAIC10-2023) and Beijing Academy of Agriculture and Forestry Sciences Postdoctoral Foundation (2022-ZZ-009).

Institutional Review Board Statement: This study used historical data from normal production records at the farm, which did not cause additional harm to laying hens, therefore did not require ethics approval.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. da Silva, M.C.; Schaefer, R.; Gava, D.; Souza, C.K.; da Silva Vaz, I., Jr.; Bastos, A.P.; Venancio, E.J. Production and application of anti-nucleoprotein IgY antibodies for influenza A virus detection in swine. *J. Immunol. Methods* **2018**, *461*, 100–105. [\[CrossRef\]](#)
2. Kaur, H.; Garg, S.; Joshi, H.; Ayaz, S.; Sharma, S.; Bhandari, M. A review: Epidemics and pandemics in human history. *Int. J. Pharma Res. Health Sci.* **2020**, *8*, 3139–3142. [\[CrossRef\]](#)
3. Transboundary Animal Diseases. Available online: <https://www.fao.org/emergencies/emergency-types/transboundary-animal-diseases/en/> (accessed on 21 July 2022).
4. Liang, W.S.; He, Y.C.; Wu, H.D.; Li, Y.T.; Shih, T.H.; Kao, G.S.; Guo, H.Y.; Chao, D.Y. Ecological factors associated with persistent circulation of multiple highly pathogenic avian influenza viruses among poultry farms in Taiwan during 2015–2017. *PLoS ONE* **2020**, *15*, e0236581. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Hay, A.J.; Skehel, J.J.; McCauley, J. Characterization of influenza virus RNA complete transcripts. *Virology* **1982**, *116*, 517–522. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Kargarfard, F.; Sami, A.; Ebrahimie, E. Knowledge discovery and sequence-based prediction of pandemic influenza using an integrated classification and association rule mining (CBA) algorithm. *J. Biomed. Inform.* **2015**, *57*, 181–188. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Kargarfard, F.; Sami, A.; Mohammadi-Dehcheshmeh, M.; Ebrahimie, E. Novel approach for identification of influenza virus host range and zoonotic transmissible sequences by determination of host-related associative positions in viral genome segments. *BMC Genomics*. **2016**, *17*, 925. [\[CrossRef\]](#)
8. Lowie, T.; Callens, J.; Maris, J.; Ribbens, S.; Pardon, B. Decision tree analysis for pathogen identification based on circumstantial factors in outbreaks of bovine respiratory disease in calves. *Prev. Vet. Med.* **2021**, *196*, 105469. [\[CrossRef\]](#)
9. Souley Kouato, B.; De Clercq, K.; Abatih, E.; Dal Pozzo, F.; King, D.P.; Thys, E.; Marichatou, H.; Saegerman, C. Review of epidemiological risk models for foot-and-mouth disease: Implications for prevention strategies with a focus on Africa. *PLoS ONE* **2018**, *13*, e0208296. [\[CrossRef\]](#)
10. Nasiri, A.; Yoder, J.; Zhao, Y.; Hawkins, S.; Prado, M.; Gan, H. Pose estimation-based lameness recognition in broiler using CNN-LSTM network. *Comput. Electron. Agric.* **2022**, *197*, 106931. [\[CrossRef\]](#)
11. Wiemken, T.L.; Kelley, R.R. Machine learning in epidemiology and health outcomes research. *Annu. Rev. Public Health* **2020**, *41*, 21–36. [\[CrossRef\]](#)

12. Punyapornwithaya, V.; Klaharn, K.; Arjkumpa, O.; Sansamur, C. Exploring the predictive capability of machine learning models in identifying foot and mouth disease outbreak occurrences in cattle farms in an endemic setting of Thailand. *Prev. Vet. Med.* **2022**, *207*, 105706. [[CrossRef](#)] [[PubMed](#)]
13. Ebrahimi, M.; Aghagolzadeh, P.; Shamabadi, N.; Tahmasebi, A.; Alsharifi, M.; Adelson, D.L.; Hemmatzadeh, F.; Ebrahimie, E. Understanding the underlying mechanism of HA-subtyping in the level of physic-chemical characteristics of protein. *PLoS ONE* **2014**, *9*, e96984. [[CrossRef](#)] [[PubMed](#)]
14. Feng, C.; Kephart, G.; Juarez-Colunga, E. Predicting COVID-19 mortality risk in Toronto, Canada: A comparison of tree-based and regression-based machine learning methods. *BMC Med. Res. Methodol.* **2021**, *21*, 267. [[CrossRef](#)] [[PubMed](#)]
15. Zheng, H.L.; An, S.Y.; Qiao, B.J.; Guan, P.; Huang, D.S.; Wu, W. A data-driven interpretable ensemble framework based on tree models for forecasting the occurrence of COVID-19 in the USA. *Environ. Sci. Pollut. Res.* **2023**, *30*, 12648–13659. [[CrossRef](#)] [[PubMed](#)]
16. Fu, Z.; Xu, F.; Zhou, Y.; Zhang, X.S. Pig-vet: A web-based expert system for pig disease diagnosis. *Expert Syst. Appl.* **2005**, *29*, 93–103. [[CrossRef](#)]
17. Zeineldin, M.M.; El-Raof, Y.M.A.; El-Attar, H.A.; Ghanem, M.M. Lung ultrasonography and computer-aided scoring system as a diagnostic aid for bovine respiratory disease in feedlot cattle. *Glob. Vet.* **2016**, *17*, 588–594. [[CrossRef](#)]
18. Katris, C. A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece. *Expert Syst. Appl.* **2021**, *166*, 114077. [[CrossRef](#)]
19. Assad, D.B.N.; Cara, J.; Ortega-Mier, M. Comparing short-term univariate and multivariate time-series forecasting models in infectious disease outbreak. *Bull. Math. Biol.* **2023**, *85*, 9. [[CrossRef](#)]
20. Wang, L.Y.; Wang, X.Y.; Chen, A.X.; Jin, X.; Che, H.L. Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model. *Healthcare* **2020**, *8*, 247. [[CrossRef](#)]
21. Wang, K.; Tian, Y.A.; Liu, S.S.; Zhang, Z.Y.; Shen, L.L.; Meng, D.Q.; Li, J. Risk factors and predictive model for dermatomyositis associated with rapidly progressive interstitial lung disease. *Pharm. Pers. Med.* **2022**, *15*, 775–783. [[CrossRef](#)]
22. Huang, L.; Liu, Y.; Huang, W.; Dong, Y.; Ma, H.; Wu, K.; Guo, A. Combining random forest and XGBoost methods in detecting early and mid-term winter wheat stripe rust using canopy level hyperspectral measurements. *Agriculture* **2022**, *12*, 74. [[CrossRef](#)]
23. Zhang, G.H.; Qu, Y.J.; Niu, Y.J.; Zhang, H.X.; Sun, Q.Q.; Liu, X.P.; Li, Y.; Zhang, H.; Liu, M.D. Difference in pathogenicity of 2 strains of avian leukosis virus subgroup J in broiler chicken. *Poult. Sci.* **2019**, *98*, 2772–2780. [[CrossRef](#)] [[PubMed](#)]
24. Pan, Q.; Liu, A.J.; Zhang, F.M.; Ling, Y.; Ou, C.B.; Hou, N.; He, C. Co-infection of broilers with *Ornithobacterium rhinotracheale* and H9N2 avian influenza virus. *BMC Vet. Res.* **2012**, *8*, 104. [[CrossRef](#)] [[PubMed](#)]
25. Linke, S.; Neubauer, K.; Dorner, M.B.; Dorner, B.G.; Pauli, G.; Schweiger, B. Generation and characterisation of monoclonal antibodies against influenza virus A, subtype H5N1. *J. Virol. Methods* **2011**, *175*, 85–94. [[CrossRef](#)]
26. Gail, M.H.; Altman, D.G.; Cadarette, S.M.; Collins, G.; Evans, S.J.; Sekula, P.; Williamson, E.; Woodward, M. Design choices for observational studies of the effect of exposure on disease incidence. *BMJ Open* **2019**, *9*, e031031. [[CrossRef](#)]
27. Chen, M.H.; Liu, Q.Y.; Chen, S.H.; Liu, Y.C.; Zhang, C.H.; Liu, R.H. XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access* **2019**, *7*, 13149–13158. [[CrossRef](#)]
28. Fan, J.L.; Wang, X.K.; Wu, L.F.; Zhou, H.M.; Zhang, F.C.; Yu, X.; Lu, X.H.; Xiang, Y.Z. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [[CrossRef](#)]
29. Omotehinwa, T.O.; Oyewola, D.O. Hyperparameter optimization of ensemble models for spam email detection. *Appl. Sci.* **2023**, *13*, 1971. [[CrossRef](#)]
30. Liu, Y.; Zhuang, Y.R.; Ji, B.Y.; Zhang, G.Q.; Rong, L.; Teng, G.H.; Wang, C.Y. Prediction of laying hen house odor concentrations using machine learning models based on small sample data. *Comput. Electron. Agric.* **2022**, *195*, 106849. [[CrossRef](#)]
31. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; pp. 153–225.
32. Xu, H.Y.; Lu, Y.X.; Li, D.; Yan, C.Y.; Jiang, Y.R.; Hu, Z.; Zhang, Z.P.; Du, R.R.; Zhao, X.L.; Zhang, Y.; et al. Probiotic mediated intestinal microbiota and improved performance, egg quality and ovarian immune function of laying hens at different laying stage. *Front. Microbiol.* **2023**, *14*, 1041072. [[CrossRef](#)]
33. Adlhoch, C.; Brouwer, A.; Kuiken, T.; Miteva, A.; Mulatti, P.; Smietanka, K.; Staubach, C.; Gogin, A.; Guajardo, I.M.; Baldinelli, F. Avian influenza overview August–November 2018. *EFSA J.* **2018**, *16*, 5573. [[CrossRef](#)]
34. Pawar, S.S.; Sajjanar, B.; Lonkar, V.D.; Kurade, N.P.; Kadam, A.S.; Nirmal, A.V.; Brahmane, M.P.; Bal, S.K. Assessing and mitigating the impact of heat stress on poultry. *Adv. Anim. Vet. Sci.* **2016**, *4*, 332–341. [[CrossRef](#)]
35. Dong, G.Y.; Cong, X.; Wang, C.M.; Wu, B.; Luo, J.; Zhang, H.; Nolte, D.L.; Deliberto, T.J.; Duan, M.X.; Ji, G.J.; et al. Reassortant H9N2 influenza viruses containing H5N1-like PB1 genes isolated from black-billed magpies in Southern China. *PLoS ONE* **2011**, *6*, e25808. [[CrossRef](#)] [[PubMed](#)]
36. Sun, Y.; Liu, J. H9N2 influenza virus in China: A cause of concern. *Protein Cell* **2015**, *6*, 18–25. [[CrossRef](#)] [[PubMed](#)]
37. Wei, Y.D.; Qi, L.; Gao, H.J.; Sun, H.L.; Pu, J.; Sun, Y.P.; Liu, J.H. Generation and protective efficacy of a cold-adapted attenuated avian H9N2 influenza vaccine. *Sci. Rep.* **2016**, *6*, 30382. [[CrossRef](#)]

38. Gu, M.; Chen, H.Z.; Li, Q.H.; Huang, J.Q.; Zhao, M.J.; Gu, X.B.; Jiang, K.J.; Wang, X.Q.; Peng, D.X.; Liu, X.F. Enzootic genotype S of H9N2 avian influenza viruses donates internal genes to emerging zoonotic influenza viruses in China. *Vet. Microbiol.* **2014**, *174*, 309–315. [[CrossRef](#)]
39. Cheung, C.L.; Vijaykrishna, D.; Smith, G.J.D.; Fan, X.H.; Zhang, J.X.; Bahl, J.; Duan, L.; Huang, K.; Tai, H.; Wang, J.; et al. Establishment of influenza A virus (H6N1) in minor poultry species in southern China. *J. Virol.* **2007**, *81*, 10402–10412. [[CrossRef](#)]
40. Bi, Y.H.; Xiao, H.X.; Chen, Q.J.; Wu, Y.; Fu, L.F.; Quan, C.S.; Wong, G.; Liu, J.; Haywood, J.; Liu, Y.X.; et al. Changes in the length of the neuraminidase stalk region impact H7N9 virulence in mice. *J. Virol.* **2015**, *90*, 2142–2149. [[CrossRef](#)]
41. Zhang, T.; Bi, Y.H.; Tian, H.A.; Li, X.W.; Liu, D.; Wu, Y.; Jin, T.; Wang, Y.; Chen, Q.J.; Chen, Z.; et al. Human infection with influenza virus A(H10N8) from live poultry markets, China, 2014. *Emerg. Infect. Dis.* **2014**, *20*, 2076–2079. [[CrossRef](#)]
42. Kim, J.A.; Cho, S.H.; Kim, H.S.; Seo, S.H. H9N2 influenza viruses isolated from poultry in Korean live bird markets continuously evolve and cause the severe clinical signs in layers. *Vet. Microbiol.* **2006**, *118*, 169–176. [[CrossRef](#)]
43. La, A.; Zhang, Q.; Cicek, N.; Coombs, K.M. Current understanding of the airborne transmission of important viral animal pathogens in spreading disease. *Biosyst. Eng.* **2022**, *224*, 92–117. [[CrossRef](#)]
44. Fuentes, S.; Gonzalez Viejo, C.; Cullen, B.; Tongson, E.; Chauhan, S.S.; Dunshea, F.R. Artificial intelligence applied to a robotic dairy farm to model milk productivity and quality based on cow data and daily environmental parameters. *Sensors* **2020**, *20*, 2975. [[CrossRef](#)] [[PubMed](#)]
45. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [[CrossRef](#)]
46. Zhu, X.Q.; Wu, X.D. Class noise vs. attribute noise: A quantitative study of their impacts. *Artif. Intell. Rev.* **2004**, *22*, 177–210. [[CrossRef](#)]
47. Ji, B.Y.; Banhazi, T.; Phillips, C.J.C.; Wang, C.Y.; Li, B.M. A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. *Biosyst. Eng.* **2022**, *216*, 186–197. [[CrossRef](#)]
48. Ebrahimi, M.; Mohammadi-Dehcheshmeh, M.; Ebrahimie, E.; Petrovski, K.R. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. *Comput. Biol. Med.* **2019**, *114*, 103456. [[CrossRef](#)]
49. Zhang, B.; Dong, X.; Hu, Y.; Jiang, X.; Li, G. Classification and prediction of spinal disease based on the SMOTE-RFE-XGBoost model. *PeerJ. Comput. Sci.* **2023**, *9*, e1280. [[CrossRef](#)]
50. Taniya. Machine Learning Algorithms: A Comparison of Different Algorithms and When to Use Them. Medium. 2018. Available online: <https://medium.com/@taniyaghosh29/machine-learning-algorithms-what-are-the-differences-9b71df4f248f> (accessed on 27 May 2018).
51. Adedirán, S.A.; Ratkowsky, D.A.; Donaghy, D.J.; Malau-Aduli, A.E.O. Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. *J. Dairy Sci.* **2012**, *95*, 5344–5356. [[CrossRef](#)]
52. Kokate, L.; Dongre, V.; Khandait, V.; Kale, S. Modeling of lactation curves for prediction of standard lactation milk yield in Marathwadi buffaloes. *Indian J. Anim. Sci.* **2019**, *89*, 909–911. [[CrossRef](#)]
53. García, M. Current and future vaccines and vaccination strategies against infectious laryngotracheitis (ILT) respiratory disease of poultry. *Vet. Microbiol.* **2017**, *206*, 157–162. [[CrossRef](#)]
54. Bell, J.G. Factors limiting production efficiency and profitability from smallholder poultry production. *World's Poult. Sci. J.* **2009**, *65*, 207–210. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.