

## Article

# Prediction of Ammonia Concentration in a Pig House Based on Machine Learning Models and Environmental Parameters

Siyi Peng <sup>1,2</sup>, Jiaming Zhu <sup>1,3,4,5</sup> , Zuohua Liu <sup>1,2,3</sup>, Bin Hu <sup>1,3,4,5</sup>, Miao Wang <sup>1,2</sup> and Shihua Pu <sup>1,3,4,5,\*</sup><sup>1</sup> Chongqing Academy of Animal Sciences, Changlong Avenue, Chongqing 402460, China<sup>2</sup> College of Animal Science and Technology, Southwest University, Chongqing 402460, China<sup>3</sup> National Center of Technology Innovation for Pigs, Chongqing 402460, China<sup>4</sup> Scientific Observation and Experiment Station of Livestock Equipment Engineering in Southwest, Ministry of Agriculture and Rural Affairs, Chongqing 402460, China<sup>5</sup> Innovation and Entrepreneurship Team for Livestock Environment Control and Equipment R&D, Chongqing 402460, China

\* Correspondence: push@cqaa.cn

**Simple Summary:** With the increased development of pig farming intensification, air quality and odor emissions in pig houses are gradually attracting attention. Among them, ammonia is considered to be an important environmental indicator of pig house. Excessive accumulation of ammonia can seriously affect the growth status of pigs and also cause a potential health risk to farm workers. Therefore, it is very important to recognize the changes of ammonia in pig houses and to discharge ammonia in time for the welfare farming of pigs. In this study, three traditional machine learning algorithms and three deep learning algorithms were selected to predict the ammonia concentration in a pig house. Based on them, important environmental parameters and promising algorithms were screened out and the algorithms were evaluated for optimization. The results of the study can provide a reference for air quality regulation in pig houses.

**Abstract:** Accurately predicting the air quality in a piggery and taking control measures in advance are important issues for pig farm production and local environmental management. In this experiment, the NH<sub>3</sub> concentration in a semi-automatic piggery was studied. First, the random forest algorithm (RF) and Pearson correlation analysis were combined to analyze the environmental parameters, and nine input schemes for the model feature parameters were identified. Three kinds of deep learning and three kinds of conventional machine learning algorithms were applied to the prediction of NH<sub>3</sub> in the piggery. Through comparative experiments, appropriate environmental parameters (CO<sub>2</sub>, H<sub>2</sub>O, P, and outdoor temperature) and superior algorithms (LSTM and RNN) were selected. On this basis, the PSO algorithm was used to optimize the hyperparameters of the algorithms, and their prediction performance was also evaluated. The results showed that the R<sup>2</sup> values of PSO-LSTM and PSO-RNN were 0.9487 and 0.9458, respectively. These models had good accuracy when predicting NH<sub>3</sub> concentration in the piggery 0.5 h, 1 h, 1.5 h, and 2 h in advance. This study can provide a reference for the prediction of air concentrations in pig house environments.

**Keywords:** ammonia concentration; machine learning; prediction models; pig house



**Citation:** Peng, S.; Zhu, J.; Liu, Z.; Hu, B.; Wang, M.; Pu, S. Prediction of Ammonia Concentration in a Pig House Based on Machine Learning Models and Environmental Parameters. *Animals* **2023**, *13*, 165. <https://doi.org/10.3390/ani13010165>

Academic Editor: Andrea Pezzuolo

Received: 17 November 2022

Revised: 17 December 2022

Accepted: 29 December 2022

Published: 31 December 2022



**Copyright:** © 2022 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 intensive and large-scale pig production, air quality and odor emissions have a negative impact on the health of the pigs, the pig farm workers, and the local environment. Ammonia (NH<sub>3</sub>) concentration is an important indicator used to evaluate the environment of a piggery. The high concentration of NH<sub>3</sub> in the piggery will affect the normal growth of pigs, resulting in decreased immunity and production performance and inducing respiratory diseases [1]. Excretion of NH<sub>3</sub> from pig houses may pose a risk of respiratory illness to pig farm workers and residents living nearby [2]. When NH<sub>3</sub> is excessively discharged

into the atmosphere, it returns to the surface through atmospheric dry and wet deposition processes, causing acidification of soil and water bodies and affecting ecosystem stability [3]. Therefore, the development of tools to assist managers in anticipating changes in  $\text{NH}_3$  concentration in a piggery will ensure that timely measures can be taken to reduce the potential stress of ammonia on human and animal health, and the level of environmental pollution, factors that are important to improve animal production, animal welfare, and environmental management.

In the past, statistical models such as least squares extensive and stepwise linear regression were developed for gas concentration prediction in aquaculture environments [4,5]. However, air pollutants in farms are mixed, complex, and usually have interaction characteristics that lead to the concentrations of air pollutants having non-linear dynamics [6]. Therefore, many statistical models in the past have poor prediction of gas pollution concentration in farms. Machine learning (ML) algorithms can deal with nonlinear interactions mathematically, and they have excellent performance in feature extraction, classification, and change prediction for big data. Machine learning has been developed rapidly in recent years [7–9]. Classical machine learning algorithms include neural networks and decision trees (DT). Based on these models, random forest (RF), extreme gradient boosting (XGBoost), backpropagation neural networks (BPNN), Elman neural networks (RNN), long short-term memory (LSTM), and other algorithms have been developed [10,11]. These algorithms have been applied to the prediction and regulation of environmental factors such as the automation of indoor air management, greenhouse gas emissions, and air pollution assessment, and have achieved good results [12,13].

Although a few researchers have constructed air prediction models for farming environments based on machine learning algorithms in recent years, the environments in farming houses vary greatly from region to region, and numerous modeling attempts and screenings are needed to achieve extensive gas concentration prediction [14,15]. For example, many pig houses have started to adopt the regulation mode (called “semi-automatic regulation” in this paper) that automatically changes the ventilation rate based on the set house temperature value. In this mode, the temperature fluctuation in the house is low, but the concentration of air pollutants in the house is often still too high in autumn and winter, and there are very few corresponding models for predicting air pollutants. In addition, as far as the modeling process is concerned, the selection of machine learning algorithms and environmental parameters in feature engineering are key aspects in determining the performance of the model, and there are very few relevant reports concerning the farming environment that can draw on how to select the underlying algorithms and environmental parameters.

In this study, we evaluated the ability of three traditional machine learning and three deep learning algorithms to predict  $\text{NH}_3$  concentration in a semi-automatically regulated pig house in combination with environmental parameters. The traditional machine learning algorithms include the classical DT, as well as support vector machine (SVM) and XGBoost, which have performed well in the past for gas prediction in farming environments [16,17]. Deep learning algorithms were chosen from the common BPNN, as well as LSTM and RNN, as these models have strong regression capabilities for time series data but are rarely employed in farming environments [18,19]. For the selection of environmental parameters, the three most concerned parameters (indoor temperature, humidity, and ventilation) in the pig house were measured, as well as the temperature and rainfall outside the house, as the latter can well reflect the changing state of the natural environment outside the house. In addition, from the response principle, indoor air pressure (P),  $\text{H}_2\text{O}$ , and  $\text{CO}_2$  may also have an effect on  $\text{NH}_3$  concentration, and these three indicators were also included in the monitoring of environmental parameters [20]. The main objectives were to evaluate the performance of LSTM, RNN, BPNN, DT, SVM, and XGBoost in predicting  $\text{NH}_3$  concentrations in semi-automatically regulated pig houses, and to identify the main environmental factors affecting  $\text{NH}_3$  concentration. On this basis, two models with strong performance in predicting  $\text{NH}_3$  concentration in semi-automatic pig houses were proposed

and optimized. This study can be a reference for future work related to gas concentration prediction in different farming modes.

## 2. Materials and Methods

### 2.1. Data Collection

This study was conducted in a fattening pig house of a pig farm in Rongchang, Chongqing. More detailed information concerning this house is given in Pu et al. [21].

Environmental data were collected from 17 September 2020, to 20 October 2020. During this period, a total of 220 pigs in the pig house were evenly distributed in 22 pens, with each pig weighing 70–90 kg. An INNOVA (model 1412I, LumaSense, Inc., USA) based on the detection principle of infrared photoacoustic spectroscopy was used to monitor and record the data of NH<sub>3</sub>, CO<sub>2</sub>, and H<sub>2</sub>O every 3 min. The HOBO (U23-001, Onset, Bourne, MA, USA) was used to monitor temperature and relative humidity, and was set to record every 5 min. The monitoring points of the above indexes were near 1.7 m in the middle of the pig house channel. Meanwhile, the ventilation volume in the piggery was regulated and recorded automatically by the intelligent system (Chongqing Dahong Machinery Co., Ltd., Chongqing, China) inside the piggery. Moreover, the temperature and rainfall data outside the house were recorded by surrounding small meteorological stations.

### 2.2. Data Preprocessing

In order to ensure the prediction performance of the model, the data collected by the equipment inside and outside the piggery and the intelligent system inside the piggery were preprocessed and analyzed. First, abnormal data processing was carried out on the environmental parameter data of the pig house using Formula (1). If the absolute value of the difference between the value and its average value was greater than three times its standard deviation, the value was replaced by the average value of the data on both sides of the value. Then, the environmental parameter data were averaged for half an hour using Formula (2). Because the dimensions of sampling equipment in the piggery were different, Equation (3) was used to normalize the data.

$$|y_n - y'| > 3\sigma \quad y_n = \frac{y_{n-1} + y_{n+1}}{2}, \quad (1)$$

$$y_h = \frac{(y_1 + y_2 + \dots + y_n)}{(30/t)}, \quad (2)$$

$$y^* = \frac{(y_n - y_{min})}{(y_{max} - y_{min})}. \quad (3)$$

Here,  $y_n$  is the collected value of a pig house sensor;  $y'$  is the mean value of the sensor data sequence;  $y_n$  is the data value after abnormal data processing;  $\sigma$  is the standard deviation of sensor data sequence;  $n$  is the data point;  $y_h$  is the value after averaging every 30 min;  $t$  is the sensor acquisition time interval;  $y_{max}$  is the maximum value of the sensor data sequence;  $y_{min}$  is the minimum value of the sensor data sequence, and  $y^*$  is the normalized value.

### 2.3. Model Construction

The construction process of the six prediction models was consistent (Figure 1), and they were all carried out in the following three steps: selecting the environmental parameters to determine the feature input scheme (2.3.1), selecting and importing potential algorithms from scikit-learn or Keras libraries using Python (2.3.2), and training the input data based on different algorithms and adjusting parameters in combination with model evaluation metrics to achieve relatively good results (2.3.3).

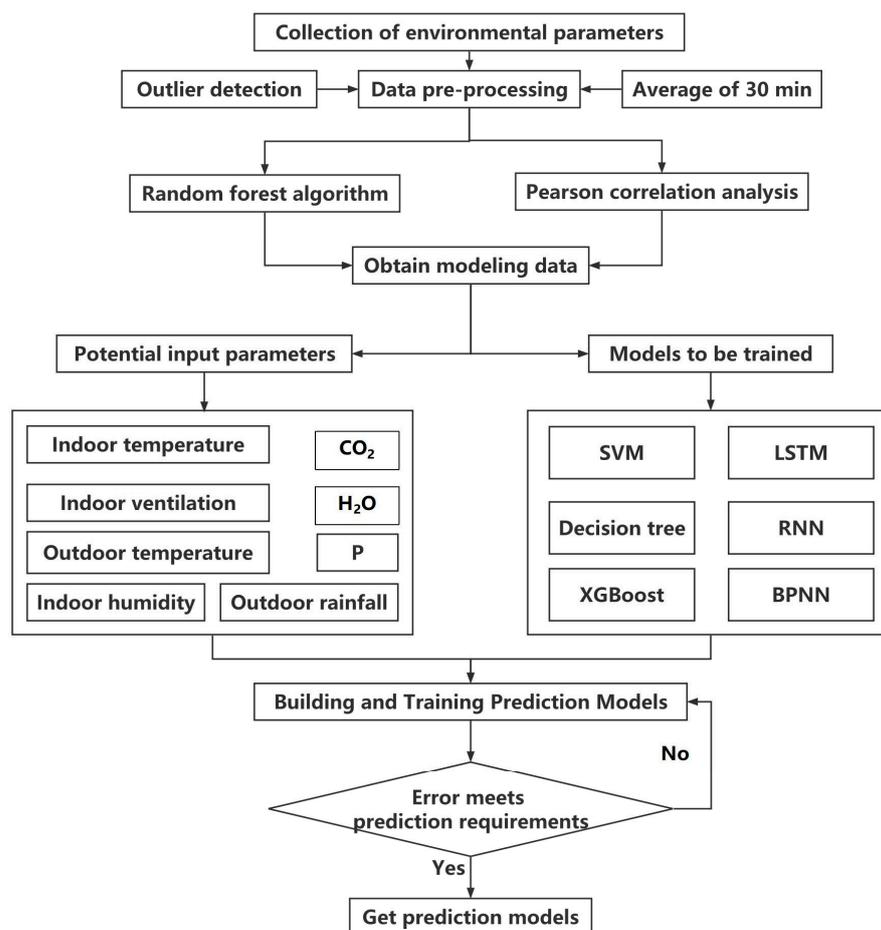


Figure 1. Modeling workflow.

### 2.3.1. Selection of Input Environmental Parameters

A variety of environmental parameters concerning the piggery were collected to build the model, including temperature, humidity, CO<sub>2</sub>, H<sub>2</sub>O, ventilation, air pressure inside the pig house, and temperature and rainfall outside the pig house. These eight parameters were considered potentially correlated variables. On this basis, the random forest algorithm was used to rank the importance of eight environmental parameters on NH<sub>3</sub> concentration in the pig house. Random forest can yield the importance score of each variable to evaluate the role of each in classification, as it relies on a self-help resampling technology and node random splitting. The ability to analyze complex interacting classification features makes random forest a feature selection tool for high-dimensional data. In this study, we considered the parameters with importance scores greater than 0.1 after random forest analysis as the priority input environmental parameters, and selected the inputs in order of importance from the largest to the smallest. The environmental parameters with importance scores less than 0.1 were used to calculate their correlations with NH<sub>3</sub> concentration using Pearson correlation analysis (PsCA), and the inputs were selected in order from the largest to the smallest according to the absolute value of correlation. The input scheme for the model characteristic parameters was obtained on the basis of the analysis of environmental importance and the correlations among the data (Table 1).

**Table 1.** Input scheme of model feature parameters.

Serial Number	Input Parameters
1	NH <sub>3</sub>
2	NH <sub>3</sub> , CO <sub>2</sub>
3	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O
4	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O, P
5	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O, P, Outdoor temperature
6	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O, P, Outdoor temperature, Indoor ventilation
7	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O, P, Outdoor temperature, Indoor ventilation, Indoor temperature
8	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O, P, Outdoor temperature, Indoor ventilation, Indoor temperature, Indoor humidity
9	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> O, P, Outdoor temperature, Indoor ventilation, Indoor temperature, Indoor humidity, Outdoor rainfall

### 2.3.2. Model Selection and Import

The NH<sub>3</sub> concentration of the pig house was used as the label datum, and the environmental parameters related to the NH<sub>3</sub> concentration were used as the characteristic data. The purpose was to learn the correspondence from the characteristic data such as temperature and humidity to predict the label data. Therefore, it was necessary to model the supervised learning algorithm in machine learning. At the same time, the input variables and output variables were time series, so the prediction of NH<sub>3</sub> in the pig house was formally a regression problem, and the corresponding model is a non-probabilistic model. Therefore, different machine learning algorithms were used to establish discriminant models in supervised learning, including classical algorithms such as neural networks, DT, SVM, and related ensemble algorithms (XGBoost, LSTM, RNN, BPNN). Using Python software, machine learning algorithm running, statistical analysis, and data mining work were managed with pandas, matplotlib, and numpy. Traditional machine learning algorithms (DT, SVM, and XGBoost) were imported directly from the scikit-learn library and combined with the input data for subsequent training and hyperparameter optimization, while deep learning algorithms (BPNN, LSTM and RNN) required additional use of the Keras library and artificial debugging to determine the number of hidden layers (there were two hidden layers in this study).

### 2.3.3. Model Training

The NH<sub>3</sub> concentration was used as the prediction target. The length of the input time series (input\_len) of each model was set to 5, and the length of the prediction time series (out\_len) was set to 1. The first 80% of the preprocessed data was used to train the model, and the last 20% was used to test the model. In the training process, the training of each integrated model involved the selection of hyperparameters, a factor that is directly related to the final prediction results. Here, the hyperparameters were firstly artificially selected and set so that the prediction effect was relatively high, and then three deep learning models and three conventional machine learning models were established. Then, the models with good prediction performance were screened, and hyperparameter optimization was performed using the corresponding algorithms on this basis. For neural network algorithms (LSTM, RNN, and BPNN), the particle swarm optimization (PSO) algorithm was used to optimize the number of hidden layer neurons in the first and second layers and the learning rate. For DT, SVM, and XGBoost algorithms, grid search was used for parameter tuning.

### 2.4. Model Performance Evaluation

The performance of the models was evaluated with mean absolute error (MAE), root-mean-squared error (RMSE), and coefficient of determination ( $R^2$ ), which are shown in Equations (4) and (5), respectively.

1. Root-Mean-Squared Error (*RMSE*)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{oi} - y_{pi})^2}. \quad (4)$$

As with *MAE*, a smaller *RMSE* means that the model prediction performance is better.

2. Coefficient of Determination ( $R^2$ )

$$R^2 = \frac{\sum_{i=1}^n (y_{pi} - y_{om})^2}{\sum_{i=1}^n (y_{pi} - y_{om})^2 + \sum_{i=1}^n (y_{oi} - y_{pi})^2}, \quad (5)$$

where  $y_{om}$  is the mean value of the observed value. An  $R^2$  closer to 1 means the model is better.

### 3. Results and Discussion

#### 3.1. Data Characteristics

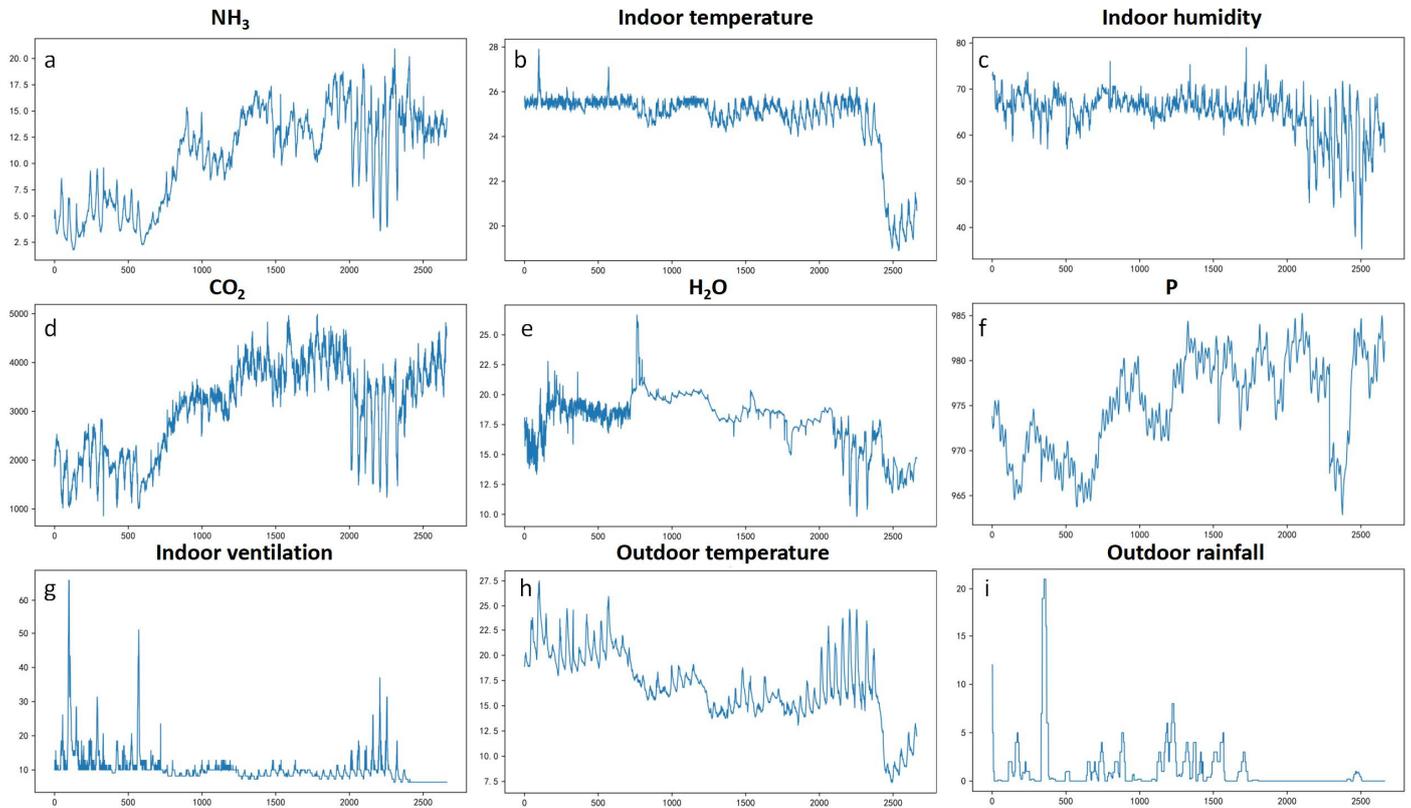
The collected parameter information is shown in Figure 2. The average concentration of  $\text{NH}_3$  fluctuated in the range of 1.77–20.94 ppm, and its association with  $\text{CO}_2$  concentration showed a significant upward trend from the 550th to the 2000th time point. Interestingly, the outdoor temperature and ventilation rate were opposite to the change trends of  $\text{NH}_3$  and  $\text{CO}_2$  concentration, fluctuating in the range of 7.5–27.0 °C and 6.0–67.5  $\text{m}^3/\text{min}$ , respectively. The temperature and humidity in the house were relatively stable in the first 2000 time points, fluctuating in the range of 23.5–28.0 °C and 57.6%–78.5%, respectively. From the 2000th to the 2300th time point, humidity in the house fluctuated significantly, and the environmental parameters near the time period changed as well, including a short-term rise in the temperature outside the house, a short-term increase in the ventilation volume in the house, and fluctuation of humidity, air pressure,  $\text{NH}_3$  concentration, and  $\text{CO}_2$  concentration in the house.

The concentration of  $\text{NH}_3$  met the standards of 25  $\text{mg}\cdot\text{m}^{-3}$ , while the  $\text{CO}_2$  did not meet the respective standard of 1500  $\text{mg}\cdot\text{m}^{-3}$  as prescribed by The Ministry of Agriculture of the People's Republic of China, given in NY/T 17824.3-2008 "Environmental parameters and environmental management for intensive pig farms." The semi-automatic control of the piggery in this experiment was able to automatically control the ventilation rate based on the temperature, so the temperature in the piggery remained relatively stable for most of the time. At the same time, when the temperature outside the house decreases, the ventilation inside the house is subsequently reduced, which in turn allows air pollutants to start accumulating in the pig house [22,23]. This is perhaps the main reason why  $\text{NH}_3$  and  $\text{CO}_2$  concentrations gradually increased after the 550th time point. It is worth noting that there was a brief increase in the outside temperature from the 2000th to 2300th time points, and the ventilation rate of the house increased automatically; this may also be the reason for the decreases in  $\text{NH}_3$  and  $\text{CO}_2$  concentrations at this time point. Thus, it seems that excessive concentrations of air pollutants in the pig house can occur, and a timely increase in ventilation in the pig house can effectively control the environment to a certain extent.

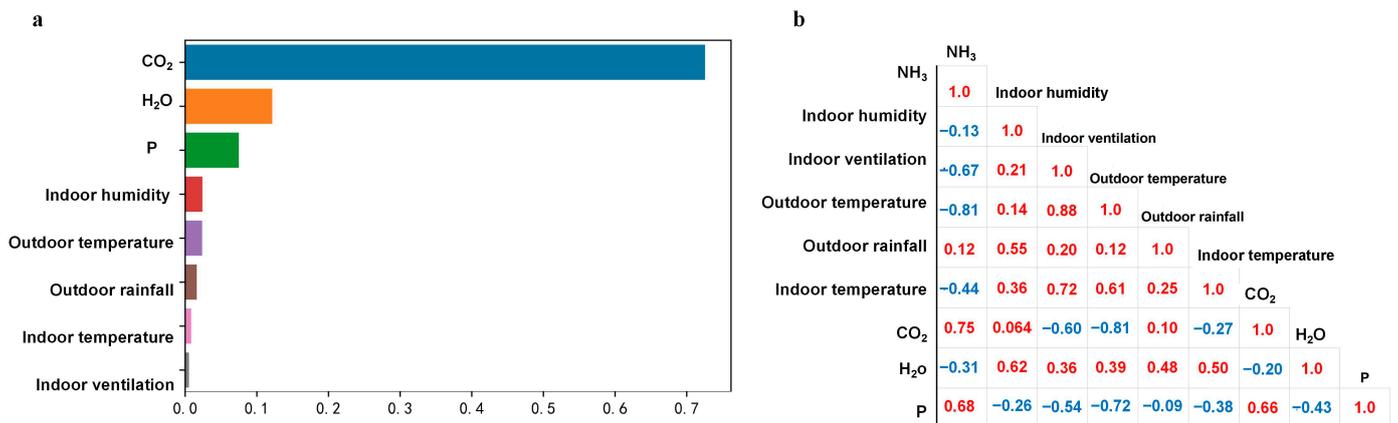
#### 3.2. Importance and Correlation of Environmental Parameters

The RF algorithm was used to evaluate the environmental variables affecting the concentrations of air pollutants in the piggery, and the importance of each variable was obtained and sorted (Figure 3a). The most important influence on  $\text{NH}_3$  concentration was  $\text{CO}_2$  concentration (importance of 0.73), followed by  $\text{H}_2\text{O}$  and P (0.12 and 0.07, respectively). Humidity, outdoor rainfall, temperature, and indoor ventilation were less important. Considering that the RF algorithm may be omitted in parameter screening, a PsCA was performed between the concentrations of gaseous pollutants in the piggery and various environmental variables (Figure 3b). The results showed that there was a strong positive correlation between  $\text{CO}_2$  and  $\text{NH}_3$  concentrations (+0.75), followed by a strong

positive correlation between P and NH<sub>3</sub> concentrations (+0.68). At the same time, there was a strong negative correlation between outdoor temperature and NH<sub>3</sub> concentration (−0.81), and there were also strong negative correlations between indoor ventilation and temperature and NH<sub>3</sub> concentration (−0.67 and −0.44, respectively).



**Figure 2.** Internal and external environment of the pig house. The above graphs show the concentration of NH<sub>3</sub> (a), CO<sub>2</sub> (d) and H<sub>2</sub>O (e) inside the house, as well as the temperature (b), humidity (c), air pressure and ventilation (f,g) inside the house, and the temperature and rainfall (h,i) outside the house, respectively.



**Figure 3.** Analysis of piggery environment variables based on random forest (a) and Pearson correlation (b).

The importance of CO<sub>2</sub> to NH<sub>3</sub> concentration may be due to the formation of CO<sub>2</sub> during NH<sub>3</sub> production. Uric acid decomposition is the main source of NH<sub>3</sub> in a pig-gery [24]. Uric acid is hydrolyzed into urea and glyoxylic acid under the action of various microorganisms, and finally urea produces NH<sub>3</sub> and CO<sub>2</sub> under the action of urease [20].

In addition, NH<sub>3</sub> emissions need to be transmitted through the liquid film layer of the air to the gas film layer, and finally enter the external atmospheric environment. This process will be accompanied by H<sub>2</sub>O volatilization, and this may be the main reason why H<sub>2</sub>O had an impact on the NH<sub>3</sub> concentration in the pig house. Random forest is a classifier established in a random manner and contains multiple decision trees [25]. Although the algorithm has been verified to effectively evaluate the contribution of environmental parameters to the indicators, there may be multiple similar decision trees in the piggery environment [26,27]. If there are several environmental parameters that are important for NH<sub>3</sub> concentration because of the same mechanism, then some of them are likely to be neglected in the random forest method. Therefore, we introduced PsCA and found that P, indoor temperature, outdoor temperature, and indoor ventilation had high correlations. P changes with the external atmospheric environment and the ventilation volume in the piggery, and this may be the reason P had strong positive correlations with the outside temperature and the ventilation in the piggery. In addition, ventilation rate is an important parameter for regulating the environment of the piggery. In this study, the ventilation rate was set to increase or decrease according to the temperature inside the piggery, and the temperature inside the piggery would change with the infiltration of the temperature outside the piggery; this may be the reason for the large negative correlations between the temperature outside the piggery, the temperature inside the piggery, the ventilation rate, and NH<sub>3</sub> concentration. In general, there were interactions among environmental parameters in the pig house.

### 3.3. Model Comparison

According to the analysis performed for the environmental parameters, nine input schemes of characteristic parameters were determined in the process of training the model (Table 1), and the accuracy of each model was evaluated with the value of  $R^2$  as the index (Table 2). Meanwhile, three cases were selected for comparative analysis without feature parameters (only input NH<sub>3</sub>), partial characteristic parameters with good prediction effect (input NH<sub>3</sub>, CO<sub>2</sub>, H<sub>2</sub>O, P, and outdoor temperature), and full characteristic parameters were selected for comparative analysis (Figure 4). In general, LSTM, RNN, and XGBoost had excellent prediction results, and even with different input features; the predicted and original values of these three models in the test set mostly overlapped, especially LSTM and RNN (Table 2 and Figure 4). DT could partially predict NH<sub>3</sub> concentration, but the difference between its predicted and original values was larger than those of the first three. BPNN had good prediction results only when suitable input features (such as input NH<sub>3</sub>, CO<sub>2</sub>, H<sub>2</sub>O, P, and outdoor temperature) were used, and it deviated from the overall performance of both SVM. When the input feature was only NH<sub>3</sub>, the LSTM, RNN, and XGBoost could mostly predict NH<sub>3</sub> (the first column in Figure 4), and most of their predictions differ from the original values only at the inflection point. When the input features were NH<sub>3</sub>, CO<sub>2</sub>, H<sub>2</sub>O, P, and outdoor temperature, the LSTM, RNN, and XGBoost models produced better prediction results than others. The predicted values of the six models were closer to the original values (the second column of Figure 4) than when only NH<sub>3</sub> was input. The predicted values of LSTM, RNN, and XGBoost coincided with the original values at most of the inflection points. When all environmental parameters were used as input features (the third column of Figure 4), even for the LSTM and RNN, the deviation of the predicted values from the original values increased at the 300th time point of the test set. The difference between predicted and original values increased for the six models compared to when only NH<sub>3</sub> was input.

**Table 2.** Evaluation of model accuracy via  $R^2$  for each of the nine feature input schemes.

Input Feature Parameters	Model Algorithm					
	LSTM	RNN	BPNN	DT	SVM	XGBoost
Serial 1	0.9239	0.9176	0.5709	0.8973	0.7240	0.9171
Serial 2	0.9297	0.9214	0.8080	0.8993	0.8948	0.9234

Table 2. Cont.

Input Feature Parameters	Model Algorithm					
	LSTM	RNN	BPNN	DT	SVM	XGBoost
Serial 3	0.9348	0.9327	0.7999	0.9060	0.8975	0.9267
Serial 4	0.9335	0.9275	0.5726	0.8977	0.8078	0.9173
Serial 5	0.9321	0.9392	0.8241	0.9067	0.9137	0.9312
Serial 6	0.9115	0.9138	0.5034	0.8949	0.6853	0.9077
Serial 7	0.9183	0.9197	0.6226	0.8872	0.7875	0.9095
Serial 8	0.8780	0.8739	0.4953	0.8652	0.7899	0.8707
Serial 9	0.9102	0.9007	0.4289	0.8683	0.7755	0.8861

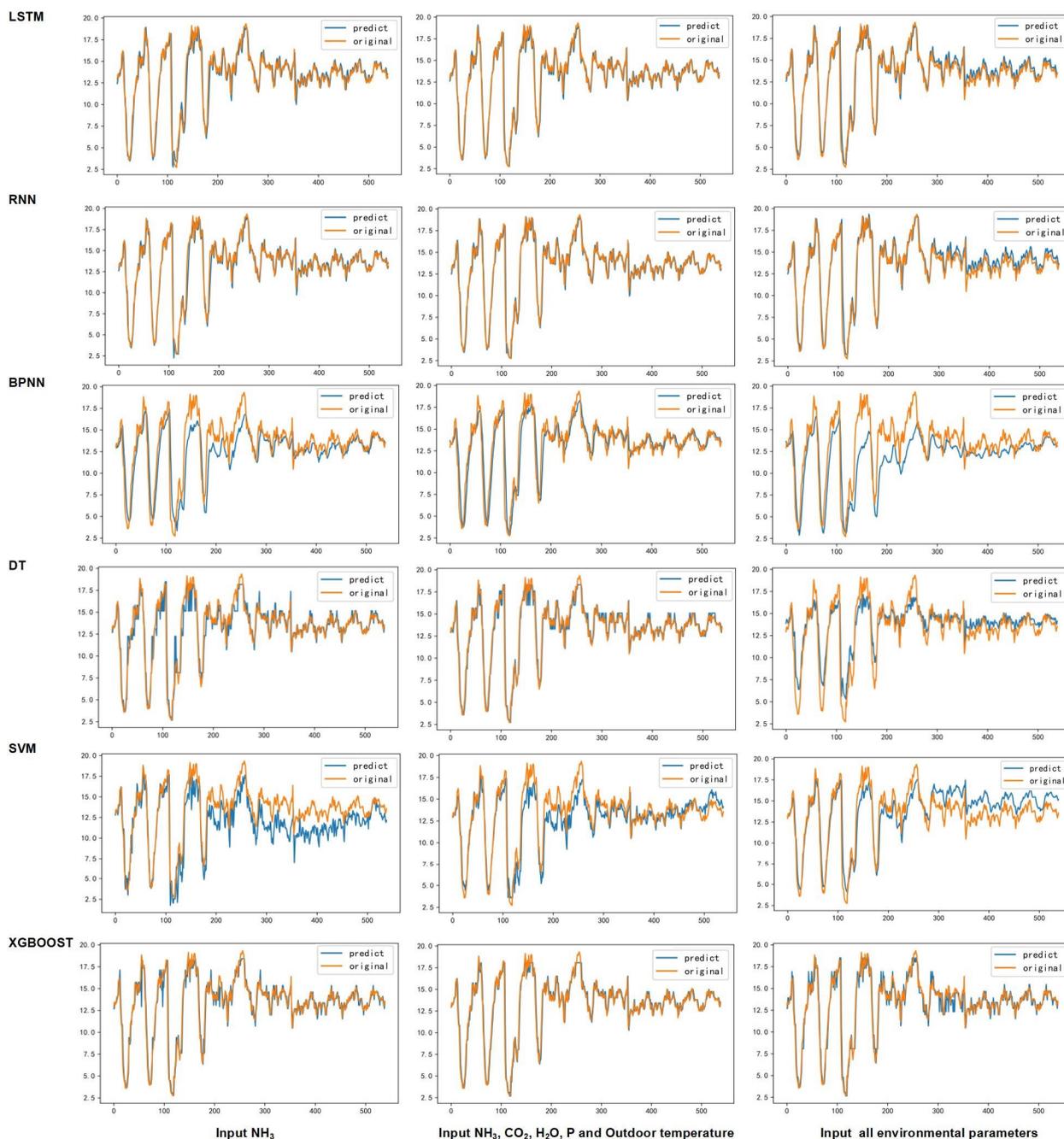


Figure 4. Comparison between predicted and original values of each model after the input of only NH<sub>3</sub> (first column), input of NH<sub>3</sub>, CO<sub>2</sub>, H<sub>2</sub>O, P, and Outdoor temperature (second column) and input of all environmental variables (third column).

LSTM and RNN have been considered as powerful algorithms for predicting atmospheric pollutant concentrations in previous studies [28,29]. XGBoost is a typical tree model for unstable classifiers that can solve nonlinear problems and has achieved good results in indoor odor prediction in the past [30,31]. In this study, when the input environmental parameters were the same, all three of the above algorithms showed strong predictive power in most cases, especially LSTM and RNN. The RNN algorithm is a kind of feedforward neural network that can transmit signals from input to output in only one way, and it introduces the self-connections of a neural cyclic structure into the network [32,33]. Therefore, the algorithm has good predictive power for data with serial characteristics. LSTM is based on RNN by introducing memory blocks to overcome vanishing and exploding gradients [34]. The memory block consists of three gating units: an input gate, an output gate, and a forget gate, where the input gate controls the flow of cell activation from the input to the memory cell, and the output gate controls the flow of output from the memory cell to other nodes [35]. Considering that both LSTM and RNN performed better than other models in this study for the nine input schemes, the results suggest that both LSTM and RNN models may have good prediction ability for NH<sub>3</sub> concentration in semi-automated pig houses.

When the input environmental parameters were altered, the  $R^2$  values of the models, even those constructed using the same algorithm, could be dramatically different. In this study, the  $R^2$  of each model with input NH<sub>3</sub>, CO<sub>2</sub>, H<sub>2</sub>O, P, and outdoor temperature were improved compared to when only NH<sub>3</sub> was input, especially for BPNN and SVM. This is consistent with previous studies that environmental parameters could increase model accuracy [36,37]. It is noteworthy that the  $R^2$  value of each model decreased when all environmental parameters were input than when only NH<sub>3</sub> was input. This could be that some of the features were not strongly correlated with changes in NH<sub>3</sub> concentration and instead negatively affected the models when they were trained [38]. In general, the input of some environmental feature parameters can improve the model accuracy, although the number of feature parameters input needs to be controlled, and suitable indicators need to be selected. For the prediction of NH<sub>3</sub> concentration in semi-automated pig houses, the characteristic parameters may firstly be considered as indicators with high importance after random forest analysis, and secondly be considered as supplementary from the perspective of correlations.

### 3.4. Model Optimization and Evaluation

Based on the analysis results of Section 3.3, LSTM and RNN models were further optimized. Here, both models comprised two hidden layers, and the number of neurons in the first and second hidden layers and the learning rate were determined by the PSO algorithm. The hyperparameters and evaluation indexes after model optimization are shown in Table 3. After optimization by PSO algorithm, both LSTM and RNN models were improved. The  $R^2$  values of PSO-LSTM and PSO-RNN increased to 0.9487 and 0.9458, respectively. In addition, LSTM and RNN were tried in combination (PSO-LSTM-RNN). The weights of the PSO-LSTM-RNN model were obtained by the optimal weighting method, and the final prediction value of the ammonia concentration in the piggery was obtained. The prediction error of the PSO-LSTM-RNN model was very close to that of PSO-LSTM and PSO-RNN, and the  $R^2$  and RMSE values of this model were 0.9416 and 0.5893, respectively.

**Table 3.** The values of hyperparameters and prediction errors of LSTM and RNN models after optimization by PSO algorithm.

Model	Hyperparameters			Prediction Errors	
	Dense1	Dense2	Learning Rate	RMSE	$R^2$
PSO-LSTM	100	259	0.001	0.5914	0.9487
PSO-RNN	100	339	0.0007	0.6125	0.9458

To further evaluate the predictive power of the optimized model, the PSO-LSTM, PSO-RNN and PSO-LSTM-RNN models were applied to the prediction at different time scales. The input length of each model was set to 15, and the output lengths were set to 1, 2, 3, 4, 5, and 6; in other words, the prediction of NH<sub>3</sub> concentration in the piggery after 0.5 h, 1 h, 1.5 h, 2 h, 2.5 h, and 3 h were realized. As seen in Table 4, all three models showed strong prediction ability for ammonia concentrations in pig houses in the next 2 h (corresponding to the next 1–4 time points), especially for 0.5 h, 1 h, and 1.5 h ( $R^2$  values > 0.93;  $RMSE$  < 0.91). With the increase in prediction time, the predicted value deviated from the actual value, and the overall prediction error became greater. When predicting the NH<sub>3</sub> concentration of the piggery after 2.5 h (corresponding to more than five time points), The  $R^2$  values for the PSO-LSTM, PSO-RNN, and PSO-LSTM-RNN models decreased below 0.9, while  $RMSE$  increased for each model.

**Table 4.** Prediction errors of PSO-LSTM and PSO-RNN models at different time scales.

Model	Time Scales	$RMSE$	$R^2$
PSO-LSTM	0.5 h	0.8626	0.9447
	1 h	0.8328	0.9382
	1.5 h	0.9105	0.9378
	2 h	1.0297	0.9182
	2.5 h	1.1729	0.8968
	3 h	1.2264	0.8773
PSO-RNN	0.5 h	0.8273	0.9433
	1 h	0.8838	0.9417
	1.5 h	0.8703	0.9353
	2 h	1.0301	0.9169
	2.5 h	1.1256	0.8856
	3 h	1.2651	0.871
PSO-LSTM-RNN	0.5 h	0.8448	0.9441
	1 h	0.8583	0.9398
	1.5 h	0.8951	0.9361
	2 h	1.0296	0.9176
	2.5 h	1.1471	0.8912
	3 h	1.2458	0.8761

The PSO algorithm, which originated from the study of social behavior of birds and fish, is an intelligent evolutionary computational method that relies on collaboration and information sharing among individuals in a population to find the optimal solution [39]. In this algorithm, each particle is a moving individual in the N-dimensional search space, and the particle has two attributes: velocity and position. A particle adjusts its position in the search space and collaborates with other particles to calculate the global optimal solution. The PSO algorithm has been widely used in the field of machine learning algorithms because of its computational simplicity and high convergence efficiency [40]. In this study, the PSO algorithm was applied to the optimization of LSTM, RNN, and LSTM-RNN, and the relatively good values for the hyperparameters of the two models were determined. The algorithm effectively improved the model accuracy.

All three models optimized by the PSO algorithm showed high accuracy ( $R^2 > 0.9$ ) in predicting 1–4 future time points; this result may indicate that these types of models have good prospects in application to the prediction of NH<sub>3</sub> concentration in pig houses at different time scales. At the same time, the accuracy of all three models was very close at all time scales, possibly due to the similarity of the LSTM and RNN algorithms [41]. However, the advantage of the higher accuracy of LSTM on long time series data was not found in this study. This may have been due to the fact that the NH<sub>3</sub> concentration in this experiment was influenced by artificial regulation from time to time, and this in turn made the pattern of NH<sub>3</sub> changes over longer times behave unpredictably. In addition, for PSO-LSTM-RNN, although the number of hidden layers was increased to three during the construction of

this model, a setting that was somewhat different from the single model, the accuracy of the model was not significantly improved after the combination of the two; however, this modification did increase the complexity of the model and the computer operation burden, which are factors that related to the similarity of the principles of RNN and LSTM algorithms. Overall, the PSO-LSTM and PSO-RNN models could effectively predict NH<sub>3</sub> concentrations in a pig house at four future time points with a balance of model complexity and accuracy, and thus they have good application prospects. In addition, if a further combination of models is needed in the future, it may be necessary to consider model construction from the perspective of synergy or complementarity between algorithms.

#### 4. Conclusions

There are complex interactions among various environmental parameters in a piggery. In this study, random forest and PsCA were used to retain important characteristic parameters as much as possible while controlling the input variables of the model. Through comparative experiments, it was found that after inputting appropriate environmental parameters (e.g., CO<sub>2</sub>, H<sub>2</sub>O, P, and outdoor temperature) the accuracy of each model for predicting ammonia concentrations was superior to that when only NH<sub>3</sub> was input, while the accuracy of each model decreased after inputting too many environmental parameters. The LSTM and RNN models were selected, which were able to effectively predict the NH<sub>3</sub> concentration in a semi-automatic pig house. On this basis, the PSO-LSTM and PSO-RNN models were proposed by using the PSO algorithm. These models were more accurate than LSTM and RNN and had a good prediction effect on NH<sub>3</sub> concentration at different time scales. The PSO-LSTM and PSO-RNN models have excellent potential for application in predicting gas concentrations in breeding environments, and the introduction of other algorithms in terms of complementarity or synergy can be considered candidates with which to build more powerful combined models.

**Author Contributions:** Conceptualization, S.P. (Shihua Pu) and S.P. (Siyi Peng); methodology, S.P. (Shihua Pu) and S.P. (Siyi Peng); software, S.P. (Siyi Peng) and J.Z.; validation, B.H. and M.W.; formal analysis, S.P. (Siyi Peng) and M.W.; investigation, J.Z. and M.W.; resources, Z.L. and S.P. (Shihua Pu); data curation, S.P. (Shihua Pu) and S.P. (Siyi Peng); writing—original draft preparation, S.P. (Siyi Peng); writing—review and editing, S.P. (Shihua Pu) and Z.L.; visualization, S.P. (Shihua Pu) and Z.L.; supervision, B.H.; project administration, S.P. (Shihua Pu) and Z.L.; funding acquisition, S.P. (Shihua Pu) and Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was financially supported by the National Key Research and Development Program of China (No. 2021YFD2000803), the Modern Agroindustry Technology Research System (CARS-35), National Center of Technology Innovation For Pigs Award and Subsidy Special Project (21610).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, upon reasonable request.

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

#### References

1. Drummond, J.G.; Curtis, S.E.; Simon, J.; Norton, H.W. Effects of aerial ammonia on growth and health of young pigs. *J. Anim. Sci.* **1980**, *50*, 1085–1091. [[CrossRef](#)]
2. Philippe, F.X.; Cabaraux, J.F.; Nicks, B. Ammonia emissions from pig houses: influencing factors and mitigation techniques, *Agric. Ecosyst. Environ.* **2011**, *141*, 245–260. [[CrossRef](#)]
3. De Schrijver, A.; Nachtergale, L.; Roskams, P.; De Keersmaecker, L.; Mussche, S.; Lust, N. Soil acidification along an ammonium deposition gradient in a Corsican Pine stand in northern Belgium. *Environ. Pollut.* **1998**, *102*, 427–431. [[CrossRef](#)]
4. Janes, K.R.; Yang, S.X.; Hacker, R.R. Single component modelling of pig farm odour with statistical methods and neural networks. *Biosyst. Eng.* **2004**, *88*, 271–279. [[CrossRef](#)]

5. Jiao, H.; Yan, T.; Wills, D.A.; Carson, A.F.; McDowell, D.A. Development of prediction models for quantification of total methane emission from enteric fermentation of young Holstein cattle at various ages, *Agric. Ecosyst. Environ.* **2014**, *183*, 160–166. [[CrossRef](#)]
6. Pan, L.L.; Yang, X.S.; DeBruyn, J. Factor analysis of downwind odours from livestock farms. *Biosyst. Eng.* **2007**, *96*, 387–397. [[CrossRef](#)]
7. Butler, K.T.; Davies, D.W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555. [[CrossRef](#)]
8. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. [[CrossRef](#)]
9. Zhu, L.T.; Chen, X.Z.; Ouyang, B.; Yan, W.C.; Lei, H.; Chen, Z.; Luo, Z.H. Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors. *Ind. Eng. Chem. Res.* **2022**, *61*, 9901–9949. [[CrossRef](#)]
10. Sayad, Y.O.; Mousannif, H.; Al Moatassime, H. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Saf. J.* **2019**, *104*, 130–146. [[CrossRef](#)]
11. Ma, J.; Ding, Y.; Cheng, J.C.P.; Jiang, F.; Tan, Y.; Gan, V.J.L.; Wan, Z. Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. *J. Clean. Prod.* **2020**, *244*, 118955. [[CrossRef](#)]
12. Gu, Y.; Li, B.; Meng, Q. Hybrid interpretable predictive machine learning model for air pollution prediction. *Neurocomputing* **2022**, *468*, 123–136. [[CrossRef](#)]
13. Barczak, R.J.; Możaryn, J.; Fisher, R.M.; Stuetz, R.M. Odour concentrations prediction based on odorants concentrations from biosolid emissions. *Environ. Res.* **2022**, *214*, 113871. [[CrossRef](#)] [[PubMed](#)]
14. Song, L.; Wang, Y.; Zhao, B.; Liu, Y.; Mei, L.; Luo, J.; Guo, X. Research on Prediction of Ammonia Concentration in QPSO-RBF Cattle House Based on KPCA Nuclear Principal Component Analysis. *Procedia Comput. Sci.* **2021**, *188*, 103–113. [[CrossRef](#)]
15. Shen, W.; Fu, X.; Wang, R.; Yin, Y.; Zhang, Y.; Singh, U.; Sun, J. A prediction model of NH<sub>3</sub> concentration for swine house in cold region based on Empirical Mode Decomposition and Elman neural network. *Inf. Process. Agric.* **2019**, *6*, 297–305. [[CrossRef](#)]
16. Liu, Y.; Zhuang, Y.; Ji, B.; Zhang, G.; Rong, L.; Teng, G.; Wang, C. Prediction of laying hen house odor concentrations using machine learning models based on small sample data. *Comput. Electron. Agric.* **2022**, *195*, 106849. [[CrossRef](#)]
17. Liu, H.X.; Li, Q.; Yan, B.; Zhang, L.; Gu, Y. Bionic electronic nose based on MOS sensors array and machine learning algorithms used for wine properties detection. *Sensors* **2019**, *19*, 45. [[CrossRef](#)]
18. Athira, V.; Geetha, P.; Vinayakumar, R.; Soman, K.P. Deepairnet: Applying recurrent networks for air quality prediction. *Procedia Comput. Sci.* **2018**, *132*, 1394–1403.
19. Navares, R.; Aznarte, J.L. Predicting air quality with deep learning lstm: Towards comprehensive models. *Ecol. Inf.* **2020**, *55*, 101019. [[CrossRef](#)]
20. Koerkamp, P. Review on emissions of ammonia from housing systems for laying hens in relation to sources, processes, building design and manure handling. *J. Agric. Eng. Res.* **1994**, *59*, 73–87. [[CrossRef](#)]
21. Pu, S.; Rong, X.; Zhu, J.; Zeng, Y.; Yue, J.; Lim, T.; Long, D. Short-Term Aerial Pollutant Concentrations in a Southwestern China Pig-Fattening House. *Atmosphere* **2021**, *12*, 103. [[CrossRef](#)]
22. Xie, Q.; Su, Z.; Ni, J.Q.; Zheng, P. Control system design and control strategy of multiple environmental factors in confined swine building. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 163–170.
23. Kim, K.Y.; Ko, H.J.; Kim, H.T.; Kim, C.N.; Byeon, S.H. Association between pig activity and environmental factors in pig confinement buildings. *Aust. J. Exp. Agric.* **2008**, *48*, 680–686. [[CrossRef](#)]
24. Ni, J. Mechanistic models of ammonia release from liquid manure: A review. *J. Agric. Eng. Res.* **1999**, *72*, 1–17. [[CrossRef](#)]
25. Thongthammachart, T.; Araki, S.; Shimadera, H.; Eto, S.; Matsuo, T.; Kondo, A. An integrated model combining random forests and WRF/CMAQ model for high accuracy spatiotemporal pm<sub>2.5</sub> predictions in the kansai region of Japan. *Atmos. Environ.* **2021**, *262*, 118620. [[CrossRef](#)]
26. Dunlop, M.W.; Blackall, P.J.; Stuetz, R.M. Odour emissions from poultry litter: A review litter properties, odour formation and odorant emissions from porous materials. *J. Environ. Manag.* **2016**, *177*, 306–319. [[CrossRef](#)]
27. Balogun, A.; Tella, A.; Baloo, L.; Adebisi, N. A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Clim.* **2021**, *40*, 100989. [[CrossRef](#)]
28. Wen, C.; Liu, S.; Yao, X.; Peng, L.; Li, X.; Hu, Y.; Chi, T. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci. Total Environ.* **2019**, *654*, 1091–1099. [[CrossRef](#)]
29. Zhao, J.; Deng, F.; Cai, Y.; Chen, J. Long short-term memory-fully connected (LSTM-FC) neural network for PM 2.5 concentration prediction. *Chemosphere* **2019**, *220*, 486–492. [[CrossRef](#)]
30. Zhi, Y.J.; Fu, D.M.; Zhang, D.W.; Yang, T.; Li, X.G. Prediction and knowledge mining of outdoor atmospheric corrosion rates of low alloy steels based on the random forests approach. *Metals* **2019**, *9*, 383. [[CrossRef](#)]
31. 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](#)]
32. Elman, J.L. Distributed representations, simple recurrent networks, and grammatical structure. *Mach. Learn.* **1991**, *7*, 195–225. [[CrossRef](#)]
33. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[PubMed](#)]
34. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]

35. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
36. Zhan, Y.; Luo, Y.; Deng, X.; Chen, H.; Grieneisen, M.L.; Shen, X.; Zhang, M. Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* **2017**, *155*, 129–139. [[CrossRef](#)]
37. Zahn, J.A.; Tung, A.E.; Roberts, B.A.; Hatfield, J.L. Abatement of ammonia and hydrogen sulphide emissions from a swine lagoon using a polymer biocover. *J. Air Waste Manag.* **2001**, *51*, 562–573. [[CrossRef](#)]
38. Choubin, B.; Abdolshahnejad, M.; Moradi, E.; Querol, X.; Mosavi, A.; Shamshirband, S.; Ghamisi, P. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci. Total Environ.* **2020**, *701*. [[CrossRef](#)]
39. Ali, J.; Chabaa, M.S.; Zeroual, A. A novel deep neural network based on randomly occurring distributed delayed PSO algorithm for monitoring the energy produced by four dual-axis solar trackers. *Renew. Energy* **2020**, *149*, 1182–1196.
40. Elmasry, W.; Akbulut, A.; Zaim, A.H. Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. *Comput. Netw.* **2020**, *168*, 107042. [[CrossRef](#)]
41. Zhang, B.; Rong, Y.; Yong, R.; Qin, D.; Li, M.; Zou, G.; Pan, J. Deep learning for air pollutant concentration prediction: A review. *Atmos. Environ.* **2022**, *290*, 119347. [[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.