Heliyon 9 (2023) e19093

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Long-term performance validation of NH_3 concentration prediction model for virtual sensor application in livestock facility

Hakjong Shin [a], Younghoon Kwak [b,*], Jung-Ho Huh [b]

[a] Department of Architectural Engineering, University of Seoul, Seoul, South Korea
[b] Department of Architecture, University of Seoul, Seoul, South Korea

ARTICLE INFO

Keywords:
Livestock
NH_3
Virtual sensor
Data-driven model
Model update

ABSTRACT

Livestock facilities commonly generate NH_3, a hazardous substance that may also harm livestock. Therefore, monitoring of NH_3 concentrations in livestock facilities is necessary to ensure proper control. However, NH_3 is alkaline and toxic, causing corrosion inside detection sensors and making monitoring difficult. This study proposes a virtual sensor concept to complement the durability of NH_3 physical sensors. The study also conducts a long-term performance validation of a data-driven NH_3 concentration prediction model. Results indicate that the model's prediction performance declines sharply when the data generation pattern inside the livestock facility changes due to changes in outdoor conditions and facility operation. Furthermore, the prediction performance of the model differed depending on the training data period settings when updating the model. Hence, the model needs versioning and update management to respond to the data generation pattern in the livestock facility when operating the NH_3 concentration virtual sensor. The virtual sensor is expected to enhance monitoring and reduce sensor management costs in livestock facilities.

1. Introduction

1.1. Background

The global population is expected to reach 8 and 10 billion by 2023 and 2056, respectively, because of the high birth rates in developing countries and the extended average life span owing to medical advancements [1]. Given this expected population increase, it is expected that the food demand in 2050 is estimated to increase by approximately 59–98% compared with that in 2016 [2]. Consequently, the increasing demand for agricultural and livestock products has led to a rise in large-scale and intensive livestock farming practices as a means to address food security issues [3–5].

In livestock facilities, livestock manure produces harmful gases, raising the concentration of these gases inside livestock facilities with high stocking densities and thereby resulting in substantially low indoor air quality [6–8]. Poor indoor air quality can lead to

Abbreviations: ANN, Artificial neural network; CART, Classification and regression tree; CFD, computational fluid dynamics; CVRMSE, coefficient of variance of the root mean square error; FDD, fault detection and diagnosis; HVAC, heating; ventilation, and air conditioning; MAE, mean absolute error; MLP, multi-layer perception; MLR, multiple linear regression; PLR, part load ratio; RBF, radial basis function; RF, random forest; SVM, support vector machine; SVR, support vector regression.

* Corresponding author.
E-mail addresses: puremorning11@uos.ac.kr (H. Shin), ikyh2@uos.ac.kr (Y. Kwak), huhj0715@uos.ac.kr (J.-H. Huh).

https://doi.org/10.1016/j.heliyon.2023.e19093
Received 19 April 2023; Received in revised form 8 August 2023; Accepted 10 August 2023
Available online 12 August 2023

respiratory and skin diseases in livestock, leading to low growth performance [9], and even affecting the health of workers in livestock facilities and nearby residents [10,11]. NH_3 is the most prevalent and harmful hazardous substance generated in livestock facilities [12]. Table 1 shows the symptoms of livestock by NH_3 concentration levels investigated in previous studies [13]. High concentrations of ammonia can result in decreased disease immunity, loss of appetite, dizziness, and other adverse effects [14–16].

Therefore, NH_3 concentrations need to be controlled according to the guidelines set [17]. However, because NH_3 is alkaline and toxic, it can cause corrosion in detection sensors [18]. Fig. 1 shows the operational history of an NH_3 sensor in a real livestock facility. During the 12-month period from September 2020 to August 2021, the livestock facility operation showed unstable ammonia measurement data, which resulted in sensor replacement five times. The considered livestock facility was a highly dense pig farm with around 900 piglets and had controlled ventilation to maintain indoor air temperature. Excluding the heating season, the ventilation fan operation schedule follows a time-series dynamics based on the outdoor condition, and accordingly, the indoor NH_3 concentration also follows time-series dynamics. However, sensor A did not show the time-series dynamics of NH_3 concentration corresponding to the ventilation fan schedule and showed a high frequency of 0 ppm or close to 0 ppm even during the period when pigs were being raised. This led to frequent sensor replacements, and even after replacing the sensor with a different company's model, sensor B, the sensor failed shortly thereafter. Finally, in mid-June 2021, sensor C was adopted and has been utilized for measurements to date with stable measurement performance. The difference in durability between these sensor models was due to the internal filter structure. In the process of replacing sensors A and B, internal corrosion was found, which may be due to the entry of NH_3, a corrosive and toxic substance. Compared to Sensor C, sensors A and B had relatively complex internal structures due to their panel mounting and were therefore vulnerable to corrosion due to insufficiently fitted filters. By contrast, Sensor C had no instrument panel and a relatively simpler internal structure with sufficient, easily replaceable filters. Certainly, Sensor C required periodic filter replacement to prevent sensor failure.

Virtual sensing technology can be an alternative to solving the practical problems of NH_3 data monitoring. Virtual sensors are a type of mathematical model applied in real-time observation of variables that are difficult to be measured physically and can be utilized for the backup of physical sensors and the replacement of faulty sensors [19]. The virtual sensing technology can be utilized for the fault detection and diagnosis (FDD) of NH_3 physical sensors as well as for deciding reasonable internal filter replacement intervals. As a result, livestock facility operators can reduce sensor installation and maintenance costs and avoid operational damage due to physical sensor malfunctions.

1.2. Literature review

Virtual sensors have been widely used in construction, transportation, healthcare, and industrial applications. In the buildings, virtual sensors are utilized to predict the indoor environment and control factors of heating, ventilation, and air conditioning (HVAC) systems, enabling optimal control and FDD of the systems [20]. This section analyzes previous studies on virtual sensor technology and NH_3 concentration prediction models for indoor building observation, ultimately suggesting the need for research related to the development of NH_3 concentration virtual sensors in livestock facilities.

1.2.1. Indoor environment virtual sensor

Virtual sensors in indoor spaces have been developed to focus on factors that are generally difficult to measure using a single sensor, such as energy consumption, thermal comfort, air infiltration, and local temperature [20]. Ploennigs et al. [21] developed a physical model-based virtual sensor that can predict room-level heat consumption by applying the concept of relative heating coefficients for room size, valve number, and heating system size factors using measured heat usage of the whole building. They believe that the newly developed virtual sensor can minimize building monitoring costs. Li et al. [22] developed an inverse model that can predict air infiltration and the number of occupants by utilizing environmental factors that are easily measurable, such as indoor temperature, humidity, and CO_2 concentration. They argued that by predicting factors that are difficult to measure, the accuracy of physics-based building energy simulation models can be improved. Zhao et al. [23] developed a Bayesian belief network-based virtual occupancy sensor utilizing a chair sensor, keyboard and mouse, real-time GPS location, and Wi-Fi connection data. They claimed that the virtual occupancy sensor can contribute to optimal control of HVAC systems by providing reliable occupancy information and can also reduce monitoring costs by utilizing only common and cheap measurements. Woradechjumroen et al. [24] developed a virtual wall surface

Table 1
Symptoms of livestock based on NH_3 concentration levels.

NH_3 concentration (ppm)	Symptoms
10	Some negative effects at long term exposure
15	Smell threshold for human beings
20	Eye irritation for broilers
20–40	Increase of respiratory diseases
25–35	Stockmen feel uncomfortable
50	Disturbance of productive capacity; Water flows from the eyes
50–150	Decrease of young pig growth by 12–29%
70	Reduced daily gain and poor feed conversion
100–200	Irritation and anorexia
5000	Deadly within a few minutes

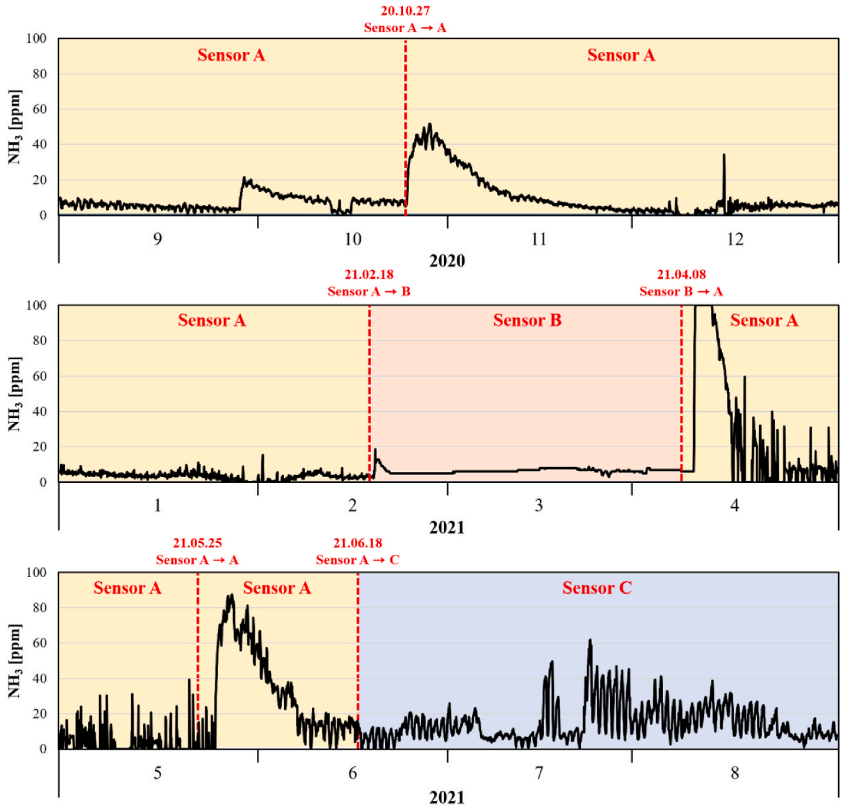

Fig. 1. Operation of NH_3 concentration monitoring sensor in a real livestock facility.

temperature sensor using a linear parametric model-based heat balance equation. They suggested that it can be used for heat transfer analysis between adjacent zones to support supervisory control of equipment in multi-zone buildings. Alhashme and Ashgriz [25] utilized computational fluid dynamics (CFD) to develop a virtual sensor that can predict the local temperature of unmeasured areas to evaluate the performance of a locally controlled temperature system.

Virtual sensors are developed with the primary goal of eliminating measurement difficulties or minimizing monitoring costs. Since NH_3 concentration sensors in livestock facilities are difficult to manage due to frequent failures caused by internal corrosion, it is difficult to properly control the indoor environment. In this situation, virtual sensing technology is expected to contribute to enhanced monitoring and cost reduction in livestock facilities by detecting NH_3 concentration sensor malfunctions and providing information such as fault diagnosis and filter replacement needs. However, research related to the development of NH_3 concentration virtual sensors is still limited because NH_3 concentration in general buildings is not monitored in the same way as in livestock facilities.

1.2.2. NH_3 concentration prediction model

No study has applied the virtual sensor concept to monitor indoor NH_3 concentration, but the development of NH_3 concentration prediction models in livestock facilities has been actively researched. Tong et al. [26] developed a 3D CFD model to predict indoor airflow, thermal environment, and NH_3 concentration distribution in a poultry farm and proposed solutions for improving the indoor environment using the developed model. Peng et al. [27] conducted a prediction performance verification by input variable combination and algorithm type when developing an data-driven NH_3 concentration prediction model in a pig farm and demonstrated the superiority of the prediction performance of the novel models developed. Song et al. [28] developed a NH_3 concentration prediction model for a cattle house using the QPSO-RBF algorithm based on data preprocessed through KPCA nuclear principal component analysis and evaluated the performance of the model in comparison with four other prediction models. Shen et al. [12] developed an NH_3 concentration prediction model for a pig farm based on the Elman neural network algorithm using environmental parametric data preprocessed by the empirical mode decomposition technique. Xie et al. [29] conducted a performance evaluation of an adaptive neuro fuzzy inference system-based NH_3 concentration prediction model for pig farms utilizing five different membership functions and derived the predicted values of NH_3 emissions in summer and winter respectively. Zhu et al. [30] developed a back propagation neural network-based NH_3 concentration prediction model for pig farms using genetic algorithm and leven berg-marquardt optimization algorithms and evaluated the prediction performance and training speed of the model.

The NH_3 concentration prediction models proposed in the previous studies can be useful for NH_3 concentration sensor management in livestock facilities because they can predict NH_3 concentrations based on the sensing values of other environmental factors. However, the previous studies are mainly aimed at improving the performance of CFD model in static situations and data-driven prediction model which is demonstrated through short-term performance validation of less than one week. However, in the actual operation stage, the performance of the prediction model may be rapidly degraded due to changes in the data generation pattern because of changes in the outdoor environment and equipment system operation in the livestock facility. Therefore, for the NH_3 concentration prediction model to play a role as a virtual sensor that can supplement physical sensors, it is necessary to conduct research related to model updates, including long-term performance validation that can reflect seasonal changes.

1.3. Research scope and goals

The study aims to:1) suggest the need for model update through long-term performance validation of a NH_3 concentration prediction model in a measured data-based livestock facility.2) suggest an update method through a cause analysis of the performance degradation of the NH_3 concentration prediction model.

This study recognizes the challenges of monitoring NH_3 concentration in livestock facilities as a problem caused by the deterioration of NH_3 physical sensors and suggests that long-term performance validation is required for NH_3 concentration prediction models to supplement physical sensors as virtual sensors. The study's view on the criteria for long-term performance validation is that the model should be able to reflect seasonal changes at a minimum. Unlike previous studies that had a validation period of less than one week, this study conducted a four-month validation of the developed model. This study is expected to strengthen monitoring in livestock facilities and reduce sensor management costs by providing necessary guidelines for the practical utilization of the NH_3 concentration prediction model as a virtual sensor.

2. Material and method

The flowchart of this study's methodology is shown in Fig. 2. First, data related to indoor temperature, relative humidity, CO_2 and NH_3 concentrations, and ventilation fan operation were collected for the development of NH_3 prediction model. Indoor NH_3 concentrations in livestock facility are primarily influenced by various factors, including the method of manure storage, type of feed, animal activity and weight, ventilation rate, indoor and outdoor environmental conditions [31]. In this study, NH_3 concentration prediction model was developed using indoor environmental factors, which are relatively easy to collect monitoring data, and fan operation schedule data directly related to ventilation. The collected data were categorized into two parts, one for training and the other for validation. Based on the data used for training, K-fold cross-validation was performed to evaluate the performance of the learning algorithm. Afterward, based on the selected algorithm, long-term performance validation of the NH_3 concentration prediction model was performed by comparing it with measured monitoring data.

2.1. Target pigsty and data monitoring

This study was conducted in a piglet house at a pig farm facility located in Suncheon, South Korea. The piglet house consisted of two piglet rooms and one diseased pig room, where approximately 900 piglets stayed for 7–10 weeks. The indoor environment of the piglet house was controlled by operating ventilation fans, cooling pads, and heating panels. The construction scale and system installations of the selected piglet house were described in detail in a previous study [32].

The experimental period was from July 2021 to December 2021. Indoor environment and part-load-ratio (PLR) of the fan were measured at 3-min intervals, as shown in Table 2. PLR represents the current rotational speed of the fan relative to its nominal rotation speed. The indoor environmental factors include air temperature, relative humidity, and CO_2 and NH_3 concentrations. The number and location of sensors installed are shown in Fig. 3. The air temperature and humidity sensors were installed at a height of 1 m from the floor. The CO_2 and NH_3 concentration sensors were installed at a height of 20 cm from the middle ceiling. An additional CO_2 concentration sensor was installed in the corridor to compare the air quality inside the pig house. The specifications of each sensor are summarized in Table 3 based on the manufacturer's catalogs.

2.2. Development of NH_3 concentration prediction model

In this study, a data-driven model was developed to predict NH_3 concentration using four parameters: indoor temperature, relative humidity, CO_2 concentration, and PLR. The measurement points of each parameter were separated and utilized as individual input parameters without any preprocessing (8 indoor temperature data points, 3 relative humidity data points, and 2 CO_2 concentration

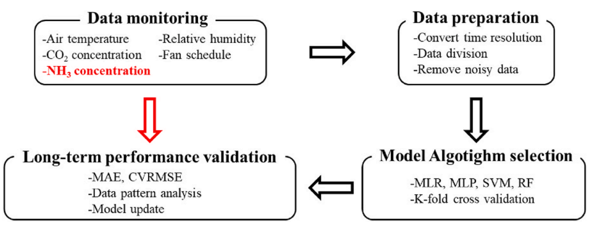

Fig. 2. Study flowchart.

Table 2
Measured data.

Term		July to December 2021
Data	Indoor environment	• Air temperature • Relative humidity • CO_2 concentration • NH_3 concentration
	Operation	• Part-load ratio (Fan)

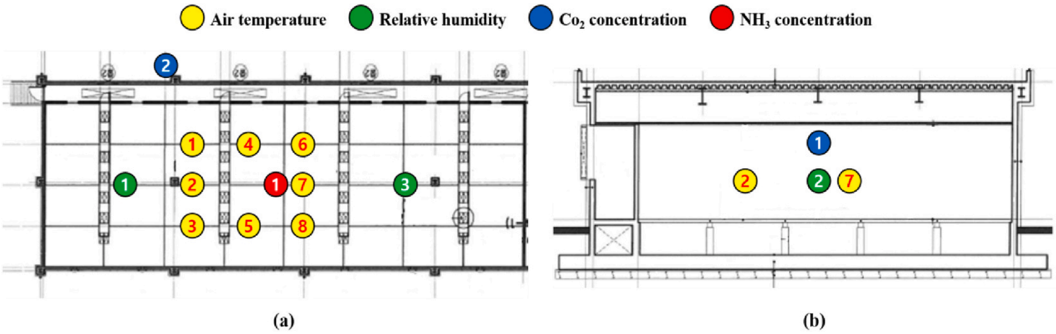

Fig. 3. Installation of sensors. (a) First-floor plan; (b) longitudinal section.

Table 3
Sensor specifications.

	Model	Manufacturer	Scope	Accuracy
Air temperature	PR-20	OMEGA	−73–260 °C	Class A per IEC60751
Relative humidity	HTX75C–W-HT	DOTECH	0-100%RH	±2.0%RH (25 °C, 20–80%RH)
CO_2 concentration	SH-VT260	SOHATECH	0–10000 ppm	±2%
NH_3 concentration	DOL53	DOL	0–100 ppm	1.5 ppm or ±10%

data points). The model was developed using the SPSS27 statistical analysis software program.

2.2.1. Data preparation

The data configuration used to train and validate the prediction model is shown in Table 4. The data collected at 9-min intervals were utilized, making a total of 7995 data points (July to August) for model development and 16,037 data points (September to December) for validating the model's performance. This data represents the entire population. Missing values and outliers were excluded without interpolation, and data collected during non-breeding periods were also excluded.

2.2.2. Data-driven model algorithm

When developing the prediction model, the performance of each algorithm was analyzed using K-fold cross-validation. K-fold cross-validation is a method that divides the dataset into K datasets, with one dataset utilized for training and (K-1) datasets for validation. Thus, a total of K dataset combinations were formed, and the model was evaluated based on the average value of K performance indicators. In this study, 5-fold cross-validation was performed for each training algorithm, as shown in Fig. 4. Long-term performance validation was performed based on the model that showed the best performance.

In this study, the performance of the NH_3 concentration prediction model was evaluated for four algorithms widely used in the field of data-driven model development: multiple linear regression (MLR), multi-layer perceptron (MLP), support vector machine (SVM), and random forest (RF). The long-term performance of the NH_3 concentration prediction model developed based on the selected

Table 4
Data division for training and Validation.

Division	Month	Number of data
Training	July	3963
	August	4032
Validation	September	3157
	October	4552
	November	4332
	December	3996

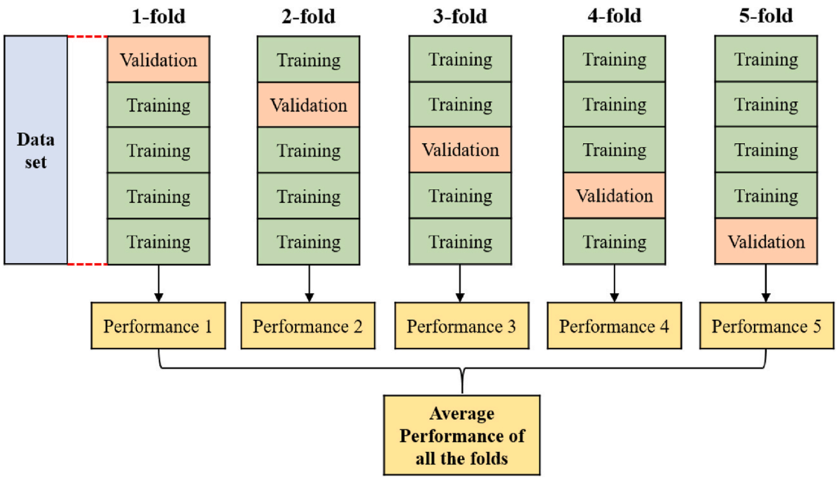

Fig. 4. Schematic diagram of 5-fold cross validation.

algorithms was evaluated.

MLR is a technique that utilizes multiple independent parameters to predict a single dependent parameter. The linear regression technique has been widely used as a prediction model in traditional statistical approaches due to the ease of model development and interpretation, and the output parameter can be derived as follows:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n \tag{1}$$

where Y is a dependent parameter corresponding to the predicted value; $X_1, X_2, X_3, \ldots X_n$ are n independent parameters; and α_0, α_1, α_2, α_3,… α_n are the regression coefficients. This equation can explain the fluctuation in Y value according to the change in values of the independent parameters and coefficients.

MLP is a type of artificial neural network (ANN) with a feed-forward approach and consists of an input layer, an output layer, and one or more hidden layers. MLPs show high performance in analyzing complex nonlinear data [33]. In general, an MLP can be represented by the following equation:

$$H_l = \sigma(w_l X + b_l) \tag{2}$$

where w is the weight; b is bias; H is the activation or output of the neurons; and σ is the activation function. This study used a sigmoid function as the activation function.

SVM creates a non-probabilistic, binary, linear classification model that determines the category of new data by utilizing a specific set of data. SVM is highly predictive even with small amounts of data [34]. SVM is an algorithm that can be used when the dependent parameter is categorical. For continuous dependent parameter prediction, Support Vector Regression (SVR) with a least-squares error function and kernel technique can be utilized [35]. SVR performs analysis with the following equation:

$$f(x) = W\hat{T}\,\varphi(x) + b \tag{3}$$

where $f(\boldsymbol{x})$ denotes the prediction outputs; $\boldsymbol{W}$ is the weight factor; b is the adjustable factor; and $\varphi(\boldsymbol{x})$ is the map function of mapping the input space into a high-dimensional feature space. In this study, the SVR algorithm with a radial basis function (RBF) kernel technique was used.

RF is an algorithm that applies the classification and regression tree (CART) algorithm to decision tree analysis and the bagging algorithm in the ensemble model. RF solves the over-fitting problem, which is a weakness of decision tree analysis, and improves the prediction accuracy by applying ensemble models. Due to these features, RF has been utilized in studies related to the development of predictive models [36,37].

2.3. Performance evaluation

To evaluate the predictive performance of the developed model, the mean absolute error (MAE) and coefficient of variance of the root mean square error (CVRMSE) metrics were used, which are often used in model validation [38]. MAE is a metric that evaluates the root mean square of the error between the observed and predicted values, as shown in Equation (4). CVRMSE is a metric that evaluates the distributional agreement by summing the squares of the errors and taking the square root again, as shown in the following equations:

$$MAE = \frac{\sum_{i=1}^{n} |M_i - P_i|}{\sum_{i=1}^{n} M_i} \times 100 \tag{4}$$

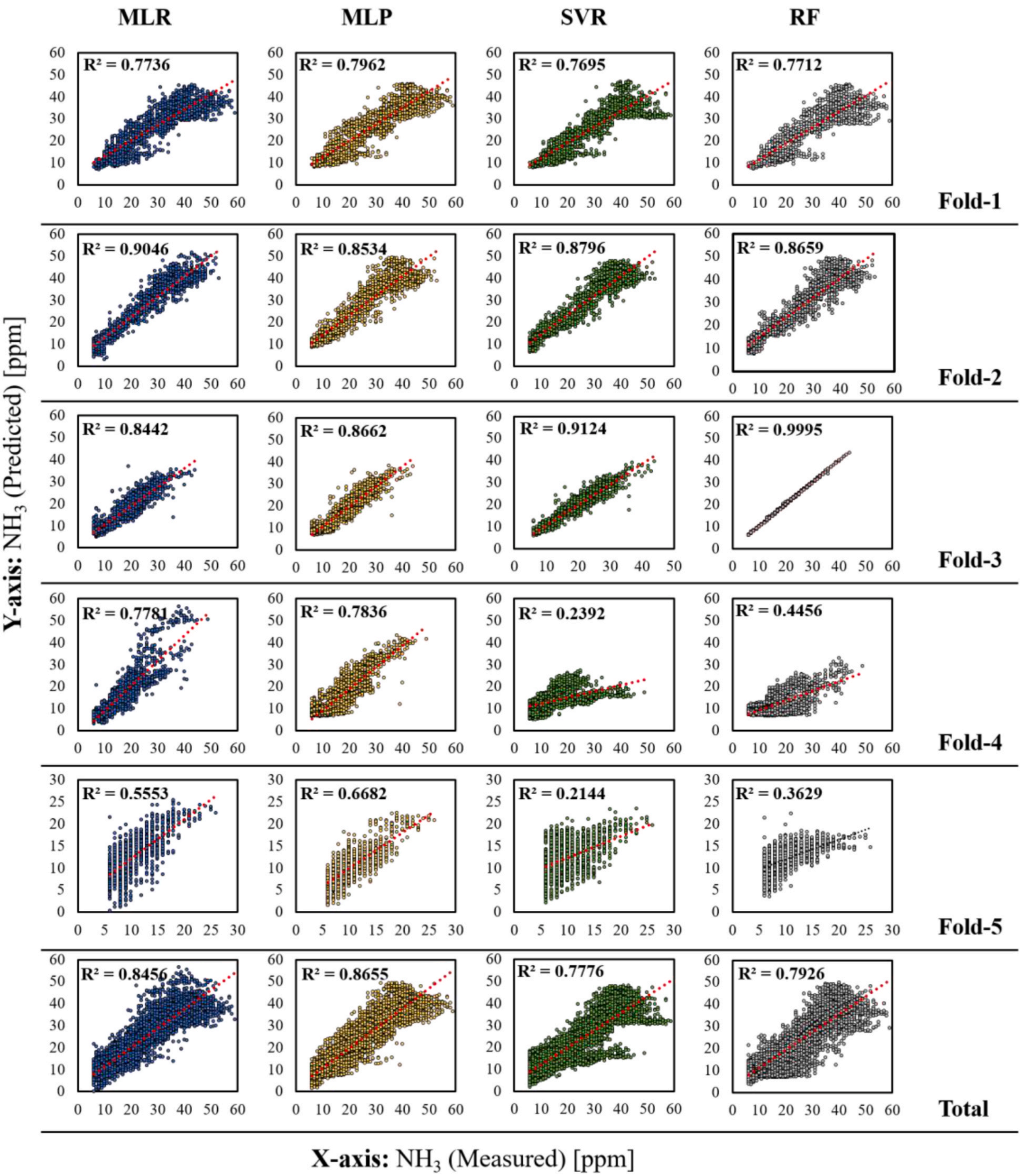

Fig. 5. Results of k-fold cross validation (Scatterplot of measured and predicted values).

$$CVRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(M_i - P_i)^2}{n}}}{\overline{M}} \times 100 \tag{5}$$

where M is the measured value; P is the predicted value; and $\overline{M}$ is the mean measured value.

3. Results and discussion

3.1. Prediction model validation and selection

In this study, the NH_3 prediction performance of each algorithm was evaluated based on July and August data from the livestock facility. Fig. 5 shows the scatterplot of predicted and measured values along with the R^2 resulting from K-fold cross-validation for each algorithm. The distribution of data near the trend line in the graph indicates the correlation between predicted and observed values. Among the four algorithms, MLP achieved the highest average R^2 value of 0.8055, while SVR had the lowest average R-squared value of 0.6321. This indicates that MLP algorithm overall predicts the observed data with the closest resemblance. The algorithm with the least variance in R-squared values across folds was MLP, with an R^2 range of 0.6682–0.8662. In contrast, SVR exhibited the largest variation in R^2 values across folds, with an R^2 range of 0.2144–0.9124. This suggests that MLP, among the four algorithms, best captures the overall characteristics of the data without overfitting to the training data. Table 5 shows the error rates for each algorithm. As a result of comparing the average values across all 5-folds, MLP had MAE = 15.67% and CVRMSE = 22.31%, among the four algorithms, with both metrics being the lowest. Based on these results, this study conducted a long-term performance validation of the NH_3 concentration prediction model developed using the MLP algorithm.

3.2. Long-term performance validation

3.2.1. Performance validation

Fig. 6 shows the measured NH_3 values and model predictions for the months of September through December. The measured NH_3 concentration in the pigsty during September–October exhibited a range of 6–55 ppm, with an average of 28.29 ppm. Similarly, the model predicted NH_3 concentration for the same period ranged from 7.02 to 54.57 ppm, with an average of 27.03 ppm. The error rates for September and October were MAE = 16.86% and CVRMSE = 21.81%. These values are similar to the results of the 5-fold validation on the July–August data that was used to develop the model. Furthermore, R^2 = 0.7137 indicates that the time series nature of the NH_3 data is well represented. However, the performance of these models began to decline sharply in November. The measured NH_3 concentration during November–December exhibited a range of 49–98 ppm, with an average of 78.17 ppm. In contrast, the model predicted NH_3 concentration for the same period ranged from 21.39 to 42.80 ppm, with an average of 33.35 ppm. The error rates for November and December were MAE = 57.33% and CVRMSE = 58.30%, indicating that the model did not perform well as a virtual sensor. The model's predictions during this period were 44.81 ppm lower than the measured average value.

This substantial degradation in the model may be attributed to changes in the pattern of data occurrence within the livestock facility. Fig. 7 shows the Pearson correlation coefficient values of NH_3 data and each input parameter by time period. The July–August data utilized for model development and the September–October data pattern where model performance remained adequate were found to be similar. Therefore, a comparison was made between the data for the four months of July–October and the November–December data. The CO_2 data was found to be positively correlated in both cases. The indoor CO_2 concentration data (CO_2_1) showed a higher correlation coefficient value in November and December compared to July and August, while the corridor CO_2 concentration data (CO_2_2) showed a lower correlation coefficient value in November and December compared to July and August. For indoor relative humidity data, the correlation coefficient was negative in July and August, while it was positive in November and December. For indoor temperature data, the sign of the correlation coefficient changed at two locations (Ta_3 and Ta_7). Overall, the absolute value of the correlation coefficient was higher in November and December compared to July and August. The PLR data

Table 5
Results of k-fold cross validation (Error rates by algorithm).

Split	Performance Index	MLR	MLP	SVR	RF
Fold 1	MAE	17.29	16.59	17.55	17.86
	CVRMSE	22.55	21.04	24.24	24.29
Fold 2	MAE	11.91	15.27	14.14	14.84
	CVRMSE	15.09	18.91	17.28	18.19
Fold 3	MAE	11.94	10.56	8.33	0.59
	CVRMSE	22.55	21.04	24.24	0.87
Fold 4	MAE	19.22	18.65	31.74	32.80
	CVRMSE	27.81	23.52	43.09	43.96
Fold 5	MAE	30.28	17.31	37.64	30.68
	CVRMSE	36.99	27.04	46.51	36.96
Average	MAE	18.13	15.67	21.88	19.36
	CVRMSE	24.99	22.31	31.07	24.86

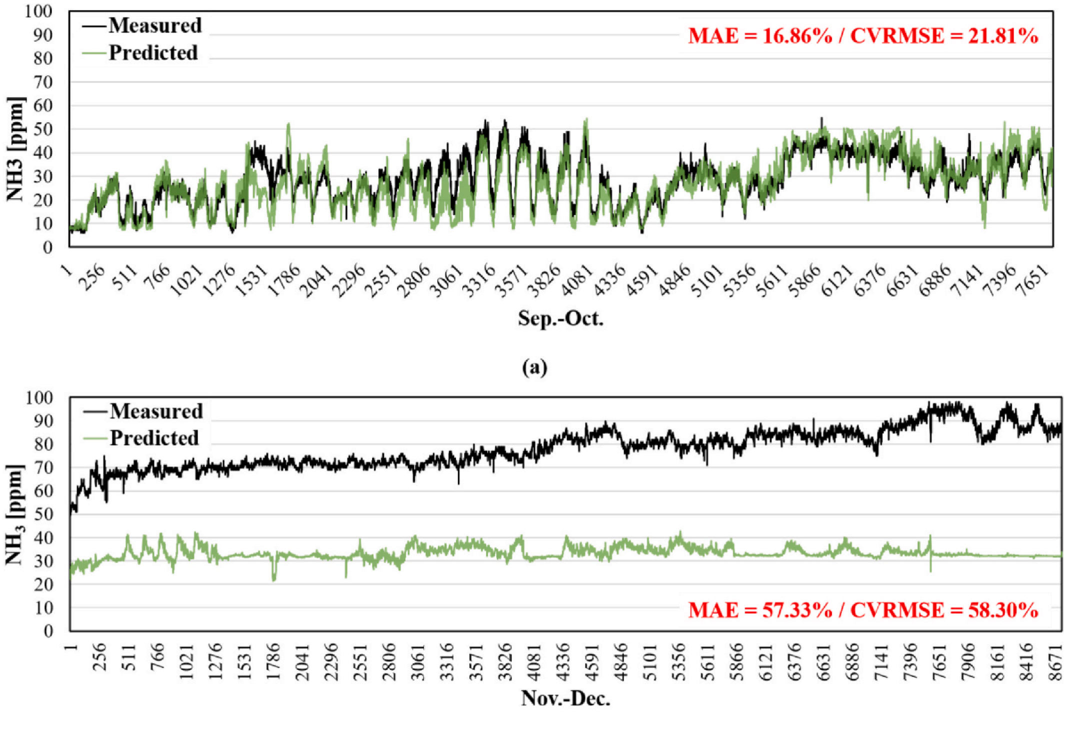

Fig. 6. Long-term performance validation of NH_3 concentration prediction model.

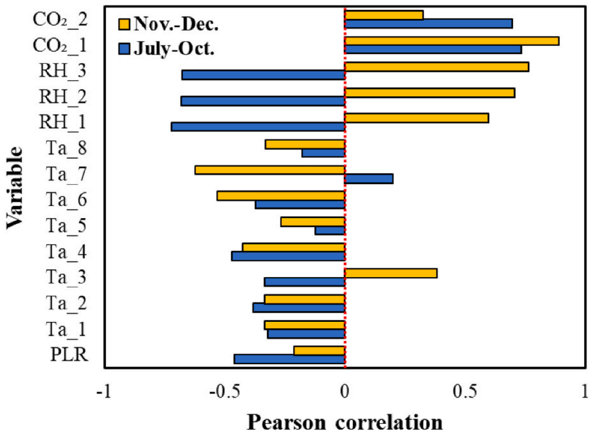

Fig. 7. Pearson correlation coefficient of each input variables for NH_3 concentration.

showed negative correlations in both periods, with a lower correlation coefficient in November–December compared to July–August.

These changes in data generation patterns were found to be attributed to seasonal variations in outdoor conditions and facility system operating conditions. Fig. 8 shows the average daily outside temperature values for July through December around the livestock facility, as published by the Korea Meteorological Administration. The graph shows an average level of 23.51 °C with no significant fluctuations until mid-October. In contrast, the temperature declined sharply thereafter, reaching freezing levels in December. In conjunction with this change in outdoor conditions, heating panels were turned on in November, which led to a steep decline in ventilation, as shown in Fig. 9. These factors resulted in different data generation patterns inside the livestock facility, which could be directly related to the degradation of the model. The observed rapid performance degradation resulting from changes in data generation patterns represents a novel finding that was not captured by short-term performance validation conducted in previous studies.

These results of the long-term performance validation suggest that updating the model periodically in response to changes in data generation patterns is crucial for a prediction model to function as a virtual sensor in a real livestock facility.

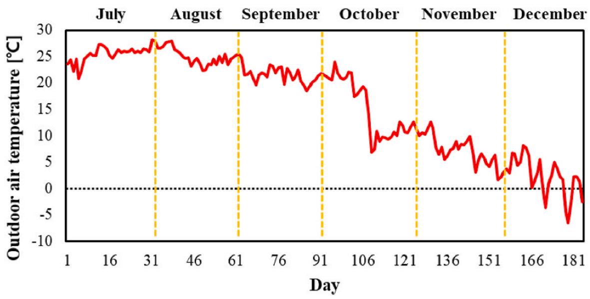

Fig. 8. Daily mean outdoor air temperature.

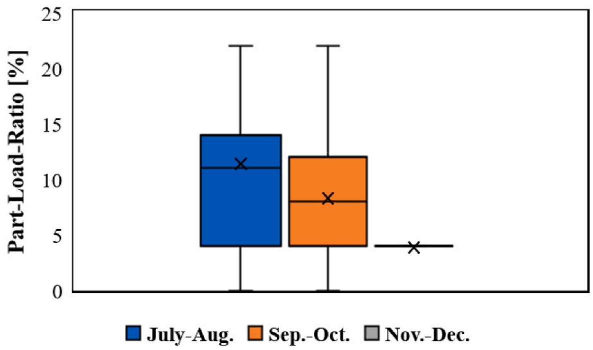

Fig. 9. Daily mean outdoor air temperature.

3.2.2. Model update

The results in Section 3.2.1 suggest that the operation of virtual sensors in livestock facilities requires a response to changes in data generation patterns due to seasonal characteristics. Therefore, in this study, besides July and August data, November data was additionally used to train the model for analyzing the change in model performance due to model replacement. To analyze whether the performance degradation of the model was caused by the insufficient amount of training data, the training data period was divided into two cases: July–November (20,036 data points) and the entire month of November (4332 data points). The two cases were then evaluated by comparing them with the measured data in December.

Fig. 10 shows the model prediction performance for the two cases mentioned above. In the case where the July–November data was used for training, the prediction performance remained at MAE = 5.07% and CVRMSE = 6.27% until the 1,108th data. However, after that, the performance declined substantially to MAE = 20.55% and CVRMSE = 21.64%. In contrast, the case utilizing data from only the month of November maintained a performance of MAE = 2.38% and CVRMSE = 3.06% for the entire month of December. This suggests that the performance degradation of the initial model in November and December was not due to the amount of training data in the model, but to the timing of the training data collection, which did not fully reflect the specific data occurrence patterns.

The changes in the performance of model, depending on the duration of the training data, suggest the need for closely analyzing the changes in data patterns, particularly those influenced by seasonal characteristics, when operating virtual sensors in livestock facilities. Moreover, it highlights the need to periodically manage the model version to keep pace with such changes.

4. Future works

4.1. Data pattern analysis

This study utilized data from July to December in a livestock facility to demonstrate the need for updating prediction model that can respond to real-time data generation patterns. Seasonal changes in the outdoor environment and facility system operations were analyzed as the cause of these data pattern changes. Therefore, it is necessary to develop a seasonal prediction model based on annual data pattern analysis. Eventually, the applicability of NH_3 virtual sensors in livestock facilities should be evaluated by verifying the predictive performance of the model through model version control.

4.2. Advancement of prediction performance

This study did not focus on improving performance of the NH_3 concentration prediction model but rather on the performance changes that may occur in real applications as a virtual sensor and conducted long-term validation. Therefore, we did not conduct a

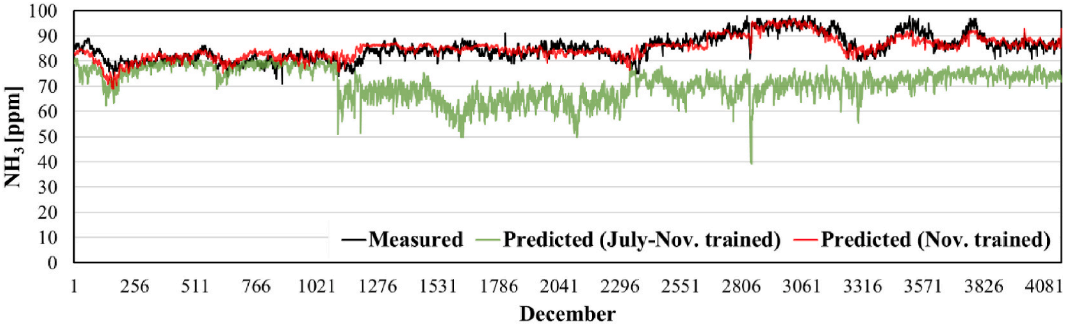

Fig. 10. Performance validation of NH_3 concentration prediction model by period of training data.

detailed study on building a development environment such as input parameter correlation analysis, data preprocessing, data interval setting, and algorithm upgrade, all of which can affect the model prediction performance. It is necessary to conduct research on the model development environment to improve the performance of the NH_3 virtual sensor in the future. Furthermore, no discussion on NH_3 concentration prediction model performance standards still exists. To actively develop and apply NH_3 concentration virtual sensors in livestock facilities, appropriate performance evaluation criteria should be prepared to reflect the physiological response of livestock and data characteristics of each concentration band.

5. Conclusions

This study conducted a long-term performance validation of a NH_3 concentration prediction model for a livestock facility. A machine learning-based NH_3 concentration prediction model was developed using measured data. Long-term performance validation was conducted for four months through comparison with measured data. The results of the study can be summarized as follows:

1) Performance analysis by algorithm through K-fold cross-validation in the model training stage showed that the MLP algorithm had the highest prediction performance with MAE = 15.67% and CVRMSE = 22.31%.
2) Long-term performance validation of the MLP-based prediction model showed that the error rates during September and October were MAE = 16.86% and CVRMSE = 21.81%. In November and December, the error rates were MAE = 57.33% and CVRMSE = 58.30%, indicating a sharp decline in performance.
3) When evaluating the prediction performance by setting the training data period at the model update stage, MAE = 20.55% and CVRMSE = 21.64% were obtained when the July–November data was used for training. In contrast, MAE = 2.38% and CVRMSE = 3.06% were obtained when data from the month of November was used for training.

Through the above research results, the necessity and method of updating the NH_3 prediction model, which is the main purpose of this study, was presented. Also, these results suggest that the practical utilization of a NH_3 concentration prediction model as a virtual sensor requires an update so as to respond to the seasonally changed data generation patterns in livestock facilities. The findings of this study, obtained by analyzing the causes of the model's prediction performance degradation by the training data period, are expected to contribute to strengthening the monitoring of livestock facilities and reducing sensor management costs.

Author contribution statement

Hakjong Shin: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. Younghoon Kwak: Conceived and designed the experiments; Analyzed and interpreted the data. Jung-Ho Huh: Contributed reagents, materials, analysis tools or data.

Data availability statement

The authors do not have permission to share data.

Additional information

No additional information is available for this paper.

Declaration of competing interest

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

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00591, Development of virtual sensor framework technology to solve sensor shading zones in digital twin environment).

References

[1] United Nations, World Population Prospects 2019: Highlights, Department of Economic and Social Affairs. Population Division, 2019.
[2] M. Elferink, F. Schierhorn, Global demand for food is rising. Can we meet it? Harv. Bus. Rev. 94 (4) (2016) 2–5.
[3] U.L.C. Baldos, T.W. Hertel, Global food security in 2050: the role of agricultural productivity and climate change, Aust. J. Agric. Resour. Econ. 58 (4) (2014) 554–570.
[4] R.C. Ilea, Intensive livestock farming: global trends, increased environmental concerns and ethical solutions, J. Agric. Environ. Ethics 22 (2) (2009) 153–167.
[5] M. Deng, M. Hou, Q. Zhang, S.D. Bellingrath-Kimura, Critical livestock densities and manure management for the typical paddy rice and corn cropping systems in an intensive livestock watershed, Japan, Agric. Syst. 177 (2020), 102722.
[6] T.M. Banhazi, P. Stott, D. Rutley, V. Blanes-Vidal, W. Pitchford, Air exchanges and indoor carbon dioxide concentration in Australian pig buildings: effect of housing and management factors, Biosyst. Eng. 110 (3) (2011) 272–279.
[7] X. Wang, X. Zhao, Y. He, K. Wang, Cough sound analysis to assess air quality in commercial weaner barns, Comput. Electron. Agric. 160 (2019) 8–13.
[8] H. Shin, Y. Kwak, S. Jo, S. Kim, J. Huh, Applicability evaluation of a demand-controlled ventilation system in livestock, Comput. Electron. Agric. 196 (2022), 106907.
[9] L.H. Yu, J.C. Liu, The influence and control of air quality in pig house in winter in north China, Modern Animal Husbandry Technology 31 (6) (2016). Issue.
[10] F.X. Philippe, J.F. Cabaraux, B. Nicks, Ammonia emissions from pig houses: influencing factors and mitigation techniques, Agric. Ecosyst. Environ. 141 (3–4) (2011) 245–260.
[11] J.H. Park, T.M. Peters, R. Altmaier, R.A. Sawvel, T.R. Anthony, Simulation of air quality and cost to ventilate swine farrowing facilities in winter, Comput. Electron. Agric. 98 (2013) 136–145.
[12] W. Shen, X. Fu, R. Wang, Y. Yin, Y. Zhang, U. Singh, B. Lkhagva, J. Sun, A prediction model of NH3 concentration for swine house in cold region based on Empirical Mode Decomposition and Elman neural network, Information Processing in Agriculture 6 (2) (2019) 297–305.
[13] J.Q. Ni, A.J. Heber, T.T. Lim, Ammonia and Hydrogen Sulfide in Swine Production, Air Quality and Livestock Farming, London, UK, 2018, pp. 69–88. CRC Press.
[14] J.G. Drummond, S.E. Curtis, J. Simon, H.W. Norton, Effects of aerial ammonia on growth and health of young pigs, J. Anim. Sci. 50 (6) (1980) 1085–1091.
[15] T.D.C. Hamilton, J.M. Roe, C.M. Hayes, A.J.F. Webster, Effects of ammonia inhalation and acetic acid pretreatment on colonization kinetics of toxigenic Pasteurella multocida within upper respiratory tracts of swine, J. Clin. Microbiol. 36 (5) (1998) 1260–1265.
[16] J.B. Jones, L.R. Burgess, A.J.F. Webster, C.M. Wathes, Behavioural responses of pigs to atmospheric ammonia in a chronic choice test, Anim. Sci. 63 (3) (1996) 437–445.
[17] S.U. Yoon, S.M. Choi, J.H. Lee, A study on the development of livestock odor (ammonia) monitoring system using ICT (information and communication technology), Agriculture 12 (1) (2022) 46.
[18] D. Kwak, Y. Lei, R. Maric, Ammonia gas sensors: a comprehensive review, Talanta 204 (2019) 713–730.
[19] J. Koo, S. Yoon, In-situ sensor virtualization and calibration in building systems, Appl. Energy 325 (2022), 119864.
[20] S. Yoon, Virtual sensing in intelligent buildings and digitalization, Autom. ConStruct. 143 (2022), 104578.
[21] J. Ploennigs, A. Ahmed, B. Hensel, P. Stack, K. Menzel, Virtual sensors for estimation of energy consumption and thermal comfort in buildings with underfloor heating, Adv. Eng. Inf. 25 (4) (2011) 688–698.
[22] H. Li, T. Hong, M. Sofos, An inverse approach to solving zone air infiltration rate and people count using indoor environmental sensor data, Energy Build. 198 (2019) 228–242.
[23] Y. Zhao, W. Zeiler, G. Boxem, T. Labeodan, Virtual occupancy sensors for real-time occupancy information in buildings, Build. Environ. 93 (2015) 9–20.
[24] D. Woradechjumroen, Y. Yu, H. Li, Virtual partition surface temperature sensor based on linear parametric model, Appl. Energy 162 (2016) 1323–1335.
[25] M. Alhashme, N. Ashgriz, A virtual thermostat for local temperature control, Energy Build. 126 (2016) 323–339.
[26] Xinjie Tong, Se-Woon Hong, L. Zhao, CFD modeling of airflow, thermal environment, and ammonia concentration distribution in a commercial manure-belt layer house with mixed ventilation systems, Comput. Electron. Agric. 162 (2019) 281–299.
[27] S. Peng, J. Zhu, Z. Liu, B. Hu, M. Wang, S. Pu, Prediction of ammonia concentration in a pig house based on machine learning models and environmental parameters, Animals 13 (1) (2023) 165.
[28] L. Song, Y. Wang, B. Zhao, Y. Liu, L. Mei, J. Luo, Z. Zuo, J. Yi, X. Guo, Research on prediction of ammonia concentration in QPSO-RBF cattle house based on KPCA nuclear principal component analysis, Procedia Computer Science 188 (2021) 103–113.
[29] Q. Xie, J. Ni, Z. Su, A prediction model of ammonia emission from a fattening pig room based on the indoor concentration using adaptive neuro fuzzy inference system, J. Hazard Mater. 325 (2017) 301–309.
[30] K.H. Zhu, S.Y. Wu, Q. Li, Prediction model for piggery ammonia concentration based on genetic algorithm and optimized BP neural network, Metall. Min. Ind. 11 (2015) 6–12.
[31] J.-Q. Ni, A.J. Heber, E.L. Cortus, T.-T. Lim, B.W. Bogan, R.H. Grant, M.T. Boehm, Assessment of ammonia emissions from swine facilities in the U.S.—application of knowledge from experimental research, Environ. Sci. Pol. 22 (2012) 25–35.
[32] H. Shin, Y. Kwak, S. Jo, S. Kim, J. Huh, Development of an optimal mechanical ventilation system control strategy based on weather forecasting data for outdoor air cooling in livestock housing, Energy 268 (2023), 126649.
[33] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renew. Sustain. Energy Rev. 73 (2017) 1104–1122.
[34] F. Lauer, G. Bloch, Incorporating prior knowledge in support vector machines for classification: a review, Neurocomputing 71 (7–9) (2008) 1578–1594.
[35] V. Vapnik, The Nature of Statistical Learning Theory, SSBM, New York, 1999.
[36] M.W. Ahmad, M. Mourshed, Y. Rezgui, Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption, Energy Build. 147 (2017) 77–89.
[37] T. Liu, Z. Tan, C. Xu, H. Chen, Z. Li, Study on deep reinforcement learning techniques for building energy consumption forecasting, Energy Build. 208 (2020), 109675.
[38] D.S.K. Karunasingha, Root mean square error or mean absolute error? Use their ratio as well, Inf. Sci. 585 (2022) 609–629.