

Article

The Development of an LSTM Model to Predict Time Series Missing Data of Air Temperature inside Fattening Pig Houses

Jun-gyu Kim ¹ , Sang-yeon Lee ¹ and In-bok Lee ^{2,3,*}

¹ Agriculture, Animal & Aquaculture Intelligence Research Center, Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Daejeon 34129, Republic of Korea

² Department of Rural Systems Engineering, Research Institute for Agriculture and Life Sciences, Global Smart Farm Convergence Major, College of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

³ Research Institute of Green Eco Engineering, Institute of Green Bio Science and Technology, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

* Correspondence: iblee@snu.ac.kr

Abstract: Because of the poor environment inside fattening pig houses due to high humidity, ammonia gas, and fine dust, it is hard to accumulate reliable long-term data using sensors. Therefore, it is necessary to conduct research for filling in the missing environmental data inside fattening pig houses. Thus, this research aimed to develop a model for predicting the missing data of the air temperature inside fattening pig houses using a long short-term memory (LSTM) model, which is one of the artificial neural networks (ANNs). Firstly, the internal and external environmental data of the fattening pig house were monitored to develop the LSTM models for data filling of the missing data and to validate the developed LSTM model. The LSTM model for data filling of the missing data was developed by learning the measured temperature inside the pig house. The LSTM model developed in this study was validated by comparing the air temperature data predicted by the LSTM model with the air temperature data measured in the fattening pig house. The LSTM model was accurate within a 3.5% error rate for the internal air temperature. Finally, the accuracy and applicability of the developed LSTM model were evaluated according to the order of learning data and the length of the missing data. In the future, for information and communication technologies (ICTs) and the convergence and application of smart farms, the LSTM models developed in this study may contribute to the accumulation of reliable long-term data at the fattening pig house.

Keywords: environmental monitoring; imputation; machine learning; pig house; recurrent neural network



Citation: Kim, J.-g.; Lee, S.-y.; Lee, I.-b. The Development of an LSTM Model to Predict Time Series Missing Data of Air Temperature inside Fattening Pig Houses. *Agriculture* **2023**, *13*, 795. <https://doi.org/10.3390/agriculture13040795>

Academic Editors: Muhammad Sultan, Redmond R. Shamshiri, Md Shamim Ahamed and Muhammad Farooq

Received: 21 February 2023

Revised: 24 March 2023

Accepted: 28 March 2023

Published: 30 March 2023



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

1. Introduction

The livestock industry is a major sector of agriculture and has been continuously growing in South Korea, reaching USD 18 billion in 2020 and constituting 40% of the total agricultural production. Additionally, the pig industry in South Korea accounts for 35% of the total livestock industry, ranking it as the largest livestock industry [1].

Traditionally, the internal environments of most livestock housings are controlled based on the measured data, such as internal air temperature and relative humidity. Here, air temperature has been generally used as a basis to regulate the exhaust fans and inlet windows in order to maintain the proper environment inside the livestock houses. Environmental control using only air temperature has a limitation since it cannot consider various factors such as humidity, ammonia concentration, and odor. In winter, the minimum ventilation for the management of the internal air temperature can cause excessive humidity, ammonia concentration, and odor. Recently, with the development of the smart farm based on information and communication technology (ICT), the algorithm for the environmental control of livestock houses has been changed from control based on an environmental

factor to integrated control based on the big data of various environmental factors. Reliable long-term data on the internal environment should be accumulated to design a precise control algorithm for livestock houses.

The expansion and automation of livestock houses have induced active research and development (R&D) and the establishment of smart farms. The global demand of smart farms continues to increase from USD 2.81 billion in 2015 to USD 4.92 billion in 2020 [2]. The monitoring of environmental data, such as temperature, humidity, and gas concentration inside the livestock houses, is critical for obtaining basic data to design environment control algorithms for an automatic control system of smart farms. Therefore, the accumulation of reliable long-term data on the environment is important because it is the basic data for environmental control and the development of the smart farm.

The long-term accumulation of monitoring data can be hindered by the reduced durability of the monitoring devices installed at livestock houses for daily measurements due to the poor internal environment with high levels of humidity, gas, and fine particulate matter [3]. In addition, the frequent failure of sensors will add repair and replacement costs for the sensor. Therefore, it is necessary to conduct research for filling in the missing data on the environments inside the livestock houses.

Several studies across different research fields have employed various techniques to estimate the missing data, either by using statistical analysis such as linear regression, autoregressive integrated moving average (ARIMA), and seasonal autoregressive integrated moving average (SARIMA) [4–9]. In the field related to agriculture, several studies have been conducted focused on the imputation of meteorological data [6,7,9,10]. Xie et al. [11] used a hybrid deep learning-driven sequential concentration transport emission model (DL-CTEM) that predicts the emissions of ammonia, carbon dioxide, and hydrogen sulfide, which are major harmful gases in pig houses, and suggested optimal ventilation control using this. Additionally, since real-time thermal environment management is important, a discrete model was used to improve the accuracy of thermal environment prediction inside the laying hen house [12]. A discrete model was developed that incorporates time-period groups (TPGs), the group buffered rolling (GBR) mechanism, and TPG factors. In the case of control based on the internal environment data in livestock house environment control, it can be said that the continuity of measured data and the interpolation of the missing data are very important.

Recently, machine learning techniques have been actively applied for the imputation of the missing data, and studies for the imputation of time series data using machine learning techniques had been also conducted in the agricultural field [13–16]. Boomgard-Zagrodnik and Brown [13] developed the random forest imputation model for missing Mesonet temperature observation. The model predicted the growing degree days (GDDs) value within an average error rate of 1.4%. Moon, Lee, and Son [15] developed the two-dimensional convolutional neural network (CNN) model for imputation of missing tabular data of several greenhouses. Song, Gao, Zhao, and Zhao [16] used the recurrent neural network (RNN) and long short-term memory (LSTM) models to fill the missing data of stem moisture data of plants. However, few studies focused on predicting the missing data of the livestock houses. In particular, since there is a lot of dust and harmful gases in livestock houses, missing data may occur frequently. Moreover, the RNN model, which is one of the ANN models, is being applied more actively because it is very important to collect time series data in agricultural facilities that manage animals and plants [17–23]. Therefore, it is very likely that the RNN model will be used as a method to solve missing data in real time in livestock facilities.

Compared to other models dealing with time series data, the LSTM models do not require the nonlinear functions to be estimated, and it has demonstrated superior performance in a wide range of sequence modeling applications [24–27]. Also, if the model uses the same number of layers, LSTM models can have a more complex structure and more parameters than gated recurrent units (GRUs) [28], which are commonly used to predict data with higher accuracy [29]. In addition, LSTM models have already shown higher

accuracy than GRU in other studies [30,31]. Accordingly, it can be seen that it is appropriate to use the LSTM model using long-term data of livestock house.

In this study, the RNN model was developed for expecting the missing data of air temperature inside the fattening pig house using the field-measured data of the internal air temperature. The internal and external environmental data of the fattening pig house were monitored during the field experiment. Based on the result of the field experiment, descriptive statistics were conducted to evaluate the environmental management of the fattening pig house. Some of the measured air temperature data inside the fattening pig house were assumed to be the missing data. The RNN model for data filling of the missing data was developed by learning the measured air temperature. The RNN model developed in this study was validated by comparing the air temperature data predicted by the RNN model with the measured data by the field experiment. Finally, the accuracy and applicability of the developed RNN model were evaluated according to the order of learning data and the length of the missing data.

2. Materials and Methods

The research flow chart for predicting missing data using the RNN model is as follows (Figure 1). First, a field experiment was conducted to monitor the internal and external environments of the fattening pig house. Through the field experiment, the air temperature, relative humidity, ventilation rate, ammonia concentration inside the fattening pig house, and the external weather were monitored in the fattening pig house. These monitoring data were used to develop the RNN models for data filling of the missing data and to validate the RNN model. Based on the result of the field experiment, descriptive statistics were conducted to evaluate the environment inside the fattening pig house. In this study, the air temperature inside the fattening pig house were initially assumed to be the missing data. The RNN model for data filling of the missing data was developed by learning the measured data in other period. The RNN model developed was validated by comparing the air temperature data predicted by the RNN model with the measured air temperature. The accuracy of the RNN model was further improved by considering the periodic parameter. Finally, the accuracy and applicability of the developed RNN model were evaluated according to the order of learning data and the length of the missing data.

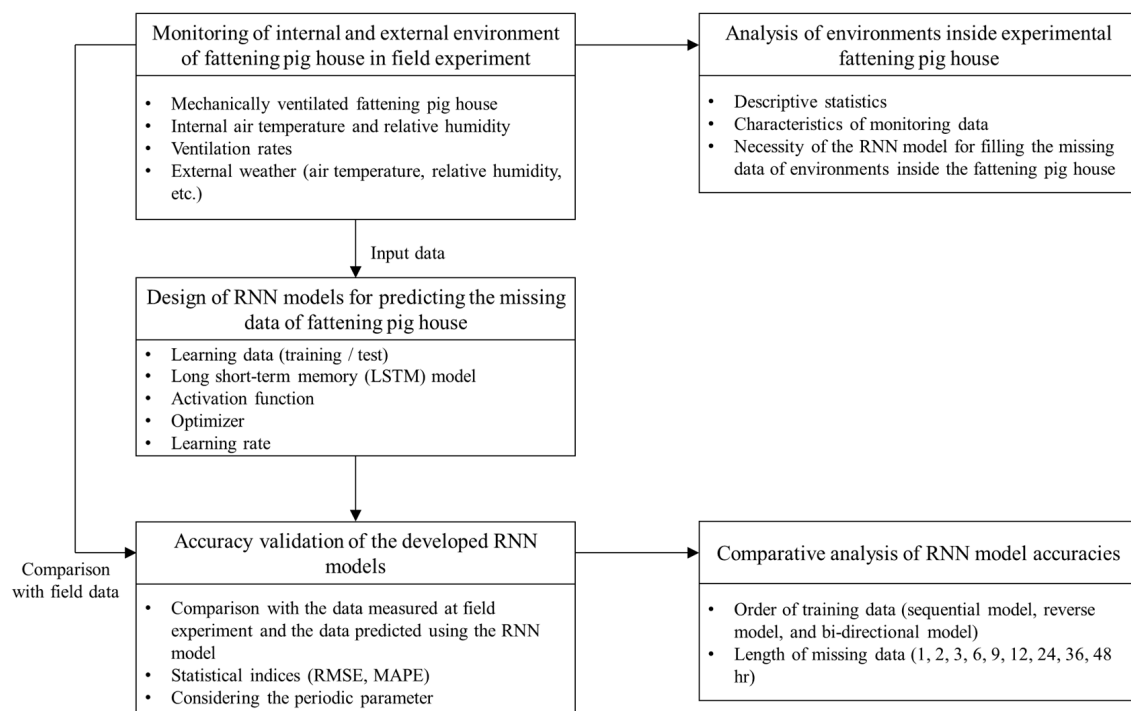


Figure 1. Flow chart of the experimental procedure of this study to develop the RNN model.

2.1. Experimental Facility (Fattening Pig House)

In this study, the experiment was conducted at the mechanically ventilated fattening pig house located at Imcheon-myeon, Buyeo-gun, Chungcheongnam-do Province ($126^{\circ}90' \text{ E}$, $36^{\circ}21' \text{ N}$). The experimental fattening pig house was shown in Figure 2. The experimental fattening pig house had a width of 42 m, a length of 145.1 m, and consisted of 24 pig rooms. The pig room where the environmental monitoring was conducted was strategically selected among the 24 pig rooms. The experimental pig room had a width of 13.3 m, a length of 18.8 m, a height of 2.6 m, and a pit depth of 1 m. The floor of the pig room was a concrete slatted floor. A total of 320 fattening pigs (about 70 kg) were reared within the experimental pig room, and the rearing density was $0.78 \text{ m}^2 \text{ animal}^{-1}$. In the experimental fattening pig house, there were six sidewall slots ($1.0 \text{ m} \times 0.4 \text{ m}$), twelve ceiling slots ($0.6 \text{ m} \times 0.6 \text{ m}$), two 0.5 m diameter exhaust fans, and three 0.95 m diameter exhaust fans for mechanical ventilation at the sidewall. The exhaust fans were controlled following the three steps for operating the exhaust fans. The first operating condition was that two 0.5 m diameter fans were operated depending on the internal air temperature. When the air temperature inside the pig house increased, the ventilation fans for temperature control were operated. The 0.95 m diameter fan started to operate when the operation rate of the first fans was 100%. Then, all the remaining fans started operating as the third operating condition when the operation rate of the second fans was 100%. In addition, the inlet ducts installed in the longitudinal direction below the ceiling have not been used.

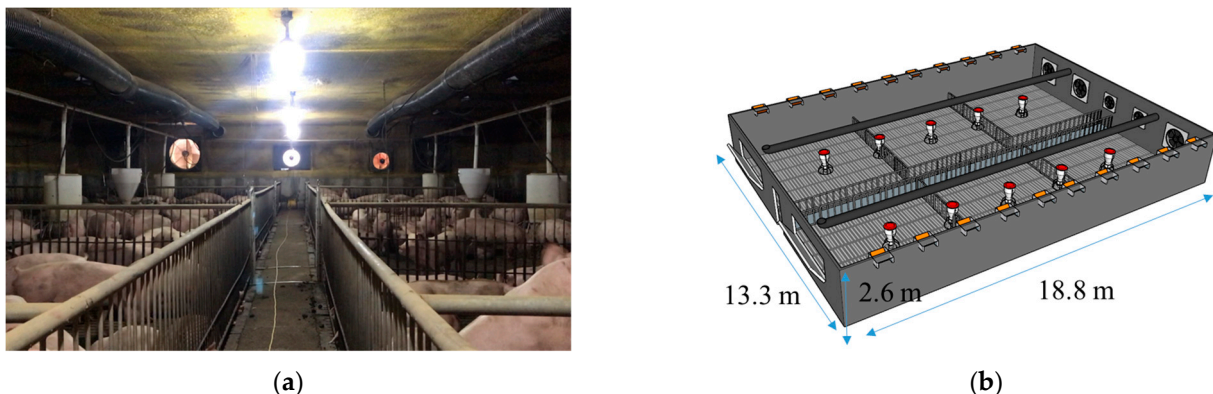


Figure 2. Experimental fattening pig house. (a) Inside the fattening pig house. (b) A schematic of the fattening pig house.

2.2. Recurrent Neural Network

As computer performance develops, machine learning technology is widely used in various fields. In previous studies, machine learning models have been used to analyze animal behavior patterns [19,20,23,32,33], behavior before calving [31,34,35], and the voice of livestock [36]. Some studies have also used machine learning to predict dependent variables according to the various environmental factors [17]. Among several machine learning technologies, artificial neural networks (ANNs) have been actively used as methods to accurately predict the dependent variables from independent variables. In this study, the RNN model, which is one of an ANN, was used to predict the missing data inside the fattening pig house. Through RNN, it is possible to use iterative learning through the memory of the ANN. The memory can store information from previous stages of learning and can provide a feedback function that considers information from previous stages as input data. The RNN structure is a form in which a path is added to the general ANN structure to re-insert the output value of the hidden layer at the previous time ($t - 1$) as the input value of the hidden layer at the next time (t). This structure repeats the process where the result at the current time (t) affects the next time ($t + 1$). A basic structure of the RNN model is shown in Figure 3a.

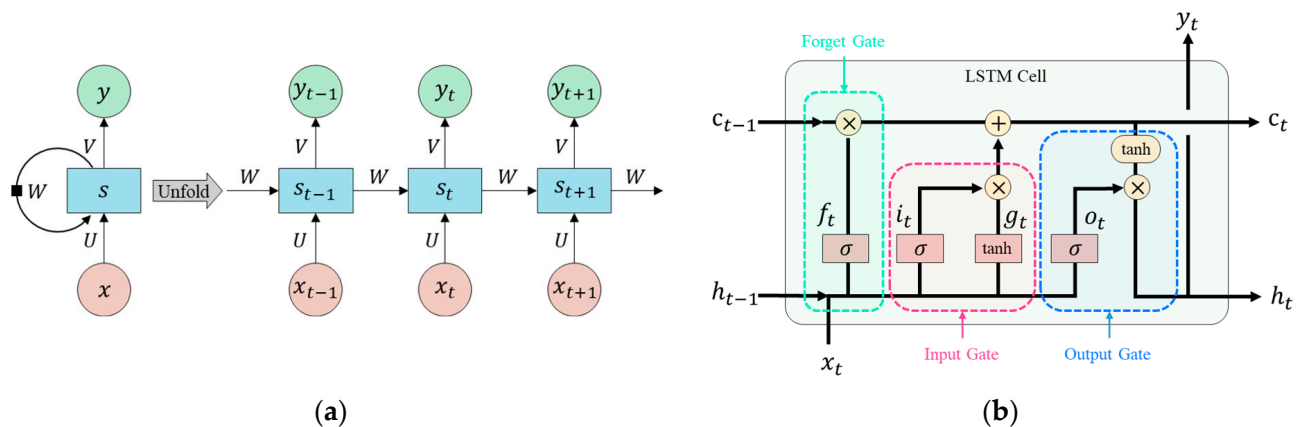


Figure 3. Basic architecture of the RNN and LSTM models [33]. (x: input data, y: output data, s: hidden data, U, V, W: weight, f_t : forget gate, i_t : input gate, o_t : output gate, h: hidden-state, c: cell-state). (a) RNN model. (b) LSTM model.

Meanwhile, the LSTM method was developed to solve the vanishing gradient problem of the general RNN algorithm [27]. A vanishing gradient is a gradient at a time step far away from time step (t) that has little effect on the learning process when learning data for a long time. While there are limitations to learning long-term dependencies using general RNN models, LSTMs can remember long previous sequences of data. The difference with the LSTM algorithm is a cell with multiple gates. LSTM accepts previous data with an additional operation, so the vanishing gradient problem does not occur. LSTM accomplishes this by using a set of gates (input, forget, and output) that control the flow of information into and out of the memory cell. The input gate can control how much new information is added to the memory cell, the forget gate controls which information is discarded from the memory cell, and the output gate controls how much information is output from the memory cell. A basic structure of the LSTM model is shown in Figure 3b.

2.3. Experimental Procedure

2.3.1. Monitoring the Environmental Factors inside the Fattening Pig House

For the data collection, the environmental factors inside and outside the experimental fattening pig house was monitored from 10 July 2021 to 31 July 2021. In this study, the monitoring data in summer was used as the learning data. The cooling system was not used at the experimental fattening pig house and the exhaust fans were maximally operated during daytime in summer. Otherwise, the internal air temperature of the fattening pig house significantly fluctuated with the ventilation, infiltration, heating, and so on. As shown in Figure 4, the six sensors (HTX 75 series, Dotech Inc., Ansan-si, Gyeonggi-do, Republic of Korea) for monitoring the air temperature and relative humidity were installed at the center of each pen at a height of 1.5 m in the fattening pig house. An ammonia sensor (Multirae-ir, RAEsystem Inc., San Jose, CA, USA) was also installed in front of the center exhaust fan. An electrometer (DW-6092, Newtech Inc., Seoul, Republic of Korea) was installed to measure the current flow when the exhaust fans were running. Then, the ventilation rates of the fattening pig house were calculated using the monitoring data of the electric current in real-time. The air temperature, relative humidity, ammonia concentration, and ventilation rate inside the fattening pig house were logged at one-second intervals. However, the data averaged for five minutes was used when the measured data were used to analyze the environments, develop the RNN model, and validate the developed RNN model. A portable weather station (Watchdog weather station 2900ET, Aurora, IL, USA) was installed on the roof of the management office near the pig house to observe the external weather conditions. External environmental data such as wind speed, wind direction, solar radiation, temperature, relative humidity, and rainfall were monitored at 1-second intervals, and the 5-minute average data was recorded on the equipment. However, the weather data

were not used as data for developing the RNN model and were used only to analyze the environment inside the pig house during the model's development process. Descriptive statistics of the monitored environments inside and outside the fattening pig house during the monitoring period were calculated for the environmental analysis of the experimental fattening pig house. The temperature humidity index (THI) [37,38] was also calculated to analyze the heat stress on the pigs. Finally, the monitoring data of air temperature inside the fattening pig house was used to develop the RNN model in this study because the air temperature is one of the most important factors for the environmental control inside the fattening pig houses. The average data of internal air temperature measured at six points was used to develop the RNN model.

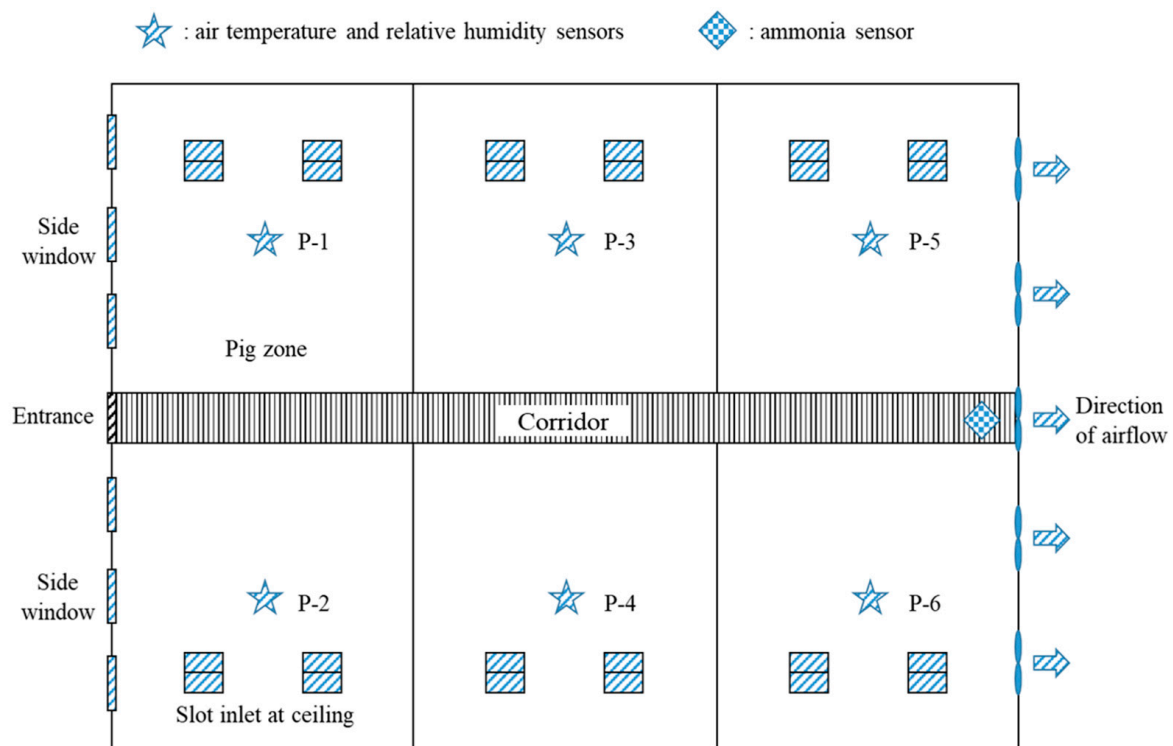


Figure 4. Sensor locations for monitoring the environmental factors inside the experimental fattening pig house.

2.3.2. Design of RNN Models for Expecting Missing Data

In this study, the RNN model was developed to predict the missing data of the air temperature from the measured data inside the fattening pig house. Since it was learned only with time series data of temperature data, it is expected that the developed model will be lightweight and highly applicable to the actual field. As previously mentioned, the measured air temperature from 10 July 2021 to 31 July 2021 was used for the RNN model's development, as shown in Figure 5. Detailed data information during the experimental period for the development and validation of the RNN model is presented in Table 1. The ratio between the lengths of training data and test data was 5:1 [16]. The data from 10 July 2021 to 19 July 2021 were used for learning in sequential order. The data house from 22 July 2021 to 31 July 2021 was used for learning in reverse order. The data house from 20 July 2021 to 21 July 2021 was used as the test data.

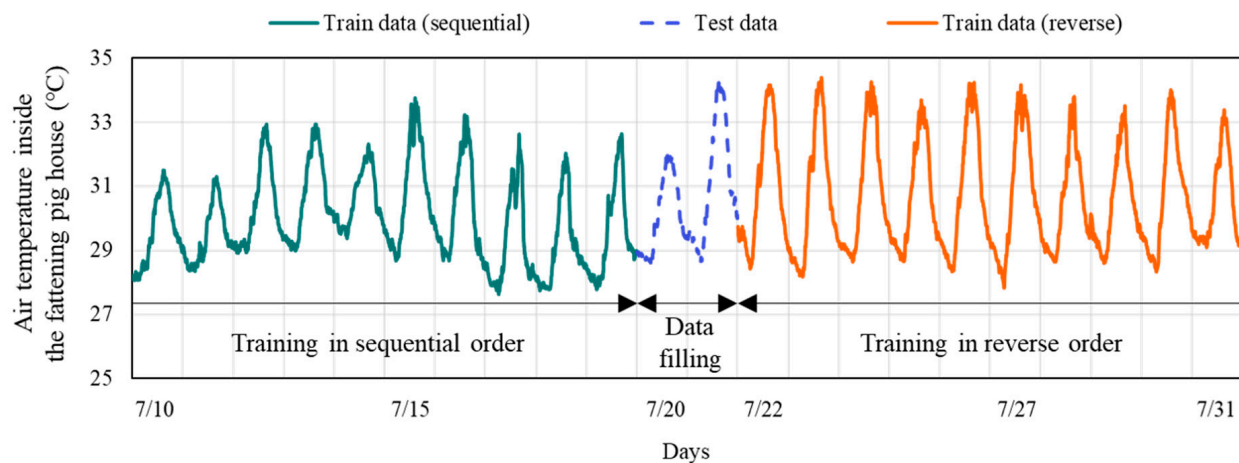


Figure 5. The train and test data set for developing the LSTM model.

Table 1. Date information periods during the experimental period.

Data Set	Monitoring Period	Days	Number of Data
Sequential data learning	10 July~19 July	10 days	2880
Validation	20 July~21 July	2 days	576
Reverse data learning	22 July~31 July	10 days	2880

Gradient vanishing problems can occur when using long-term data as training data for general RNN models. To solve this problem, a single-layer LSTM model suitable for long-term data learning was used. The learning rate was set to 0.01 and the tanh function with high accuracy of the RNN model was used as the activation function as the learning parameter. The commonly used AdamOptimizer was applied in this study, and the loss was trained to minimize the mean squared error. Since the range of learning variables can be different, the range of data according to several variables must be unified from 0 to 1. If the data range is not unified, the model diverges during the training process. Therefore, for successful learning, all training data were normalized in the range of 0 to 1 using the min–max scaler in Equation (1).

$$d_{scaled} = \frac{d - d_{min}}{(d_{max} - d_{min}) + 10^{-7}} \quad (1)$$

where d is learning data, d_{scaled} is scaled learning data, d_{max} is the maximum value of the variable, d_{min} is the minimum value of the variable, and 10^{-7} is the noise term for preventing zero division.

2.3.3. Validation of Accuracy of Developed Models

The developed RNN model was validated by comparing the temperature data predicted with the missing data assumption of the temperature data measured in the pig house. First, Model 1 was developed by learning only air temperature data inside the fattening pig house as a basic model. To improve the accuracy of the RNN model, Model 2 was developed by additionally considering a periodic parameter. The periodic parameter was set from 1 to 288 by dividing a day into five minutes to consider the periodic character of a day. The accuracies of the RNN models were compared by considering the periodic parameter (Table 2).

Table 2. Experimental conditions of learning data for developing the RNN model.

Conditions		Conditions	Number of Cases
Validation of RNN model accuracy	Learning data (independent variables)	Model 1 Monitored air temperature inside the fattening pig house	2
		Model 2 Monitored air temperature inside the fattening pig house and periodic parameter	
	Dependent variable	Missing data of air temperature	1
Total		-	2

To compare the predicted data and the measured data, statistical indices, such as the root-mean-square error (RMSE) and the mean absolute percentage error (MAPE), were calculated using Equations (2) and (3), respectively. The RMSE is commonly used to measure the difference between two pieces of data. However, there is no quantitative criterion for evaluating RMSE, and MAPE is a measure of prediction accuracy as a percentage of error. Therefore, this index was used to assess the accuracy of predicted data using the RNN model.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (R_i - C_i)^2}{n}} \quad (2)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{R_i - C_i}{R_i} \right| \quad (3)$$

where RMSE is root-mean-square error (°C, %), MAPE is mean absolute percentage error (%), n is the total number of data according to time, R_i is the measured data at a specific time, and C_i is the predicted data at a specific time.

2.3.4. Comparative Evaluation of the RNN Model according to the Order of Learning Data and the Length of the Missing Data

A total of 27 cases of experimental conditions for learning data was used to develop the RNN model, as shown in Table 3. Previous studies have shown that when developing RNN models, training time-series data in reverse order can generally improve the accuracy of RNN models [39–41]. Therefore, in this study, the RNN model was developed by learning time series data in sequential and reverse orders. The bidirectional model was developed by combining the sequential model and reverse model for improving the accuracy of the RNN models. In general, the longer the length of the missing data, the lower the accuracy of the RNN model. Therefore, to identify this trend and analyze the accuracy of the RNN model, the accuracies of the RNN models were compared according to the length of the missing data at 1, 2, 3, 6, 9, 12, 24, 36, and 48 h.

The predicted results of the bidirectional model were calculated by adding the predicted results of the sequential model and reverse model considering the weight values, as shown in Equation (4). The weight values for the predicted results of the sequential model and reverse model have gradients, as shown in Figure 6. It was expected that the sequential model has high accuracy on the front of the missing data. Otherwise, it was expected that the reverse model has high accuracy on the back of the missing data. Finally, it was expected that the accuracy of the bidirectional model is improved by comparing it with the sequential and reverse models. Finally, the accuracies of the RNN models were developed and compared according to the sequential, reverse, and bidirectional models.

$$z_i = a_i x_i + b_i y_i \quad (4)$$

where z_i is the predictive value of the bidirectional model, x_i is the predictive value of the sequential model, y_i is the predictive value of the reverse model, a_i is the weight of x_i , b_i is the weight of y_i , and i is the sequence number of the missing data.

Table 3. Experimental conditions of learning data for developing the RNN model.

Conditions	Conditions	Number of Cases
Case study for enhancing the accuracy of the RNN model	Learning data (independent variables)	Monitored air temperature inside the fattening pig house and periodic parameter
	Dependent variable	Missing data of air temperature
	Order of learning data	Sequential, reverse, and bidirectional order
	The length of the missing data	1, 2, 3, 6, 9, 12, 24, 36, and 48 h
Total	-	27

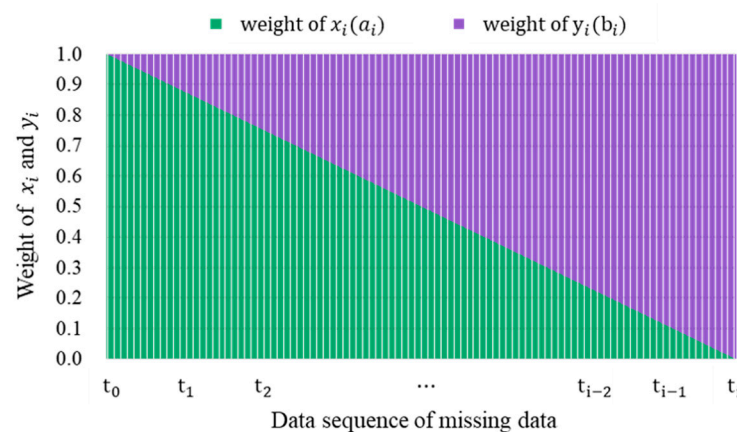


Figure 6. Weight of x_i and y_i according to the data sequence of the missing data.

3. Results and Discussions

3.1. Analysis of the Internal Environment of the Experimental Pig House

Environmental data inside and outside the experimental fattening pig house were monitored to analyze the environmental problems of the fattening pig house, to accumulate the learning data for the development of the RNN models, and to validate the developed RNN model. Descriptive statistical analysis was conducted to analyze the temperature and relative humidity characteristics inside the pig house. The box plots shown in Figure 7 describe the distributions of the air temperature and relative humidity inside the experimental fattening pig house. Figure 8 shows the measured environments inside and outside the fattening pig house during the experimental period. The average air temperature inside the fattening pig house was higher than the outside temperature due to solar radiation, insulation of the wall, heat generation of the pigs, etc. The internal air temperature was higher than 34 °C during daytime and the external air temperature was higher than 36 °C. It was expected that pigs had high-temperature stress because of the high air temperature inside the fattening pig house during the experimental period. The relative humidity inside the fattening pig house fluctuated from 50 to 90% due to the temperature difference between day and night. According to the measurement points, the analysis showed that relatively low air temperatures (28.7 °C and 29.0 °C) were measured near the slot opening at the sidewall (P-1 and P-2), whereas relatively high air temperatures (30.6 °C and 30.0 °C) were measured at P-5 and P-6, which were located near the exhaust fans. The airflow from the slot opening at the sidewall (inlets) to the exhaust fans (outlets) caused the distribution of the internal airflow.

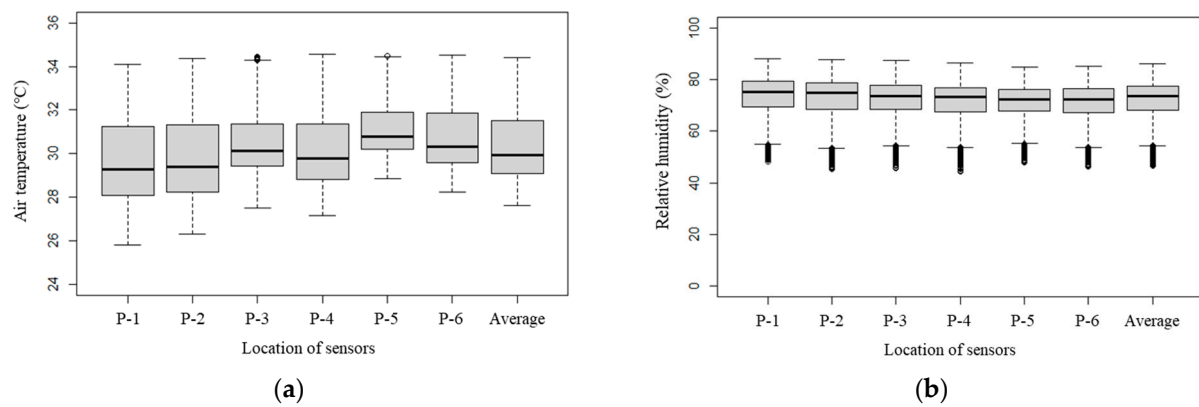


Figure 7. A box plot with the 1.5 IQR (Interquartile range) of the measured environments inside the fattening pig house. (a) Air temperature. (b) Relative humidity.

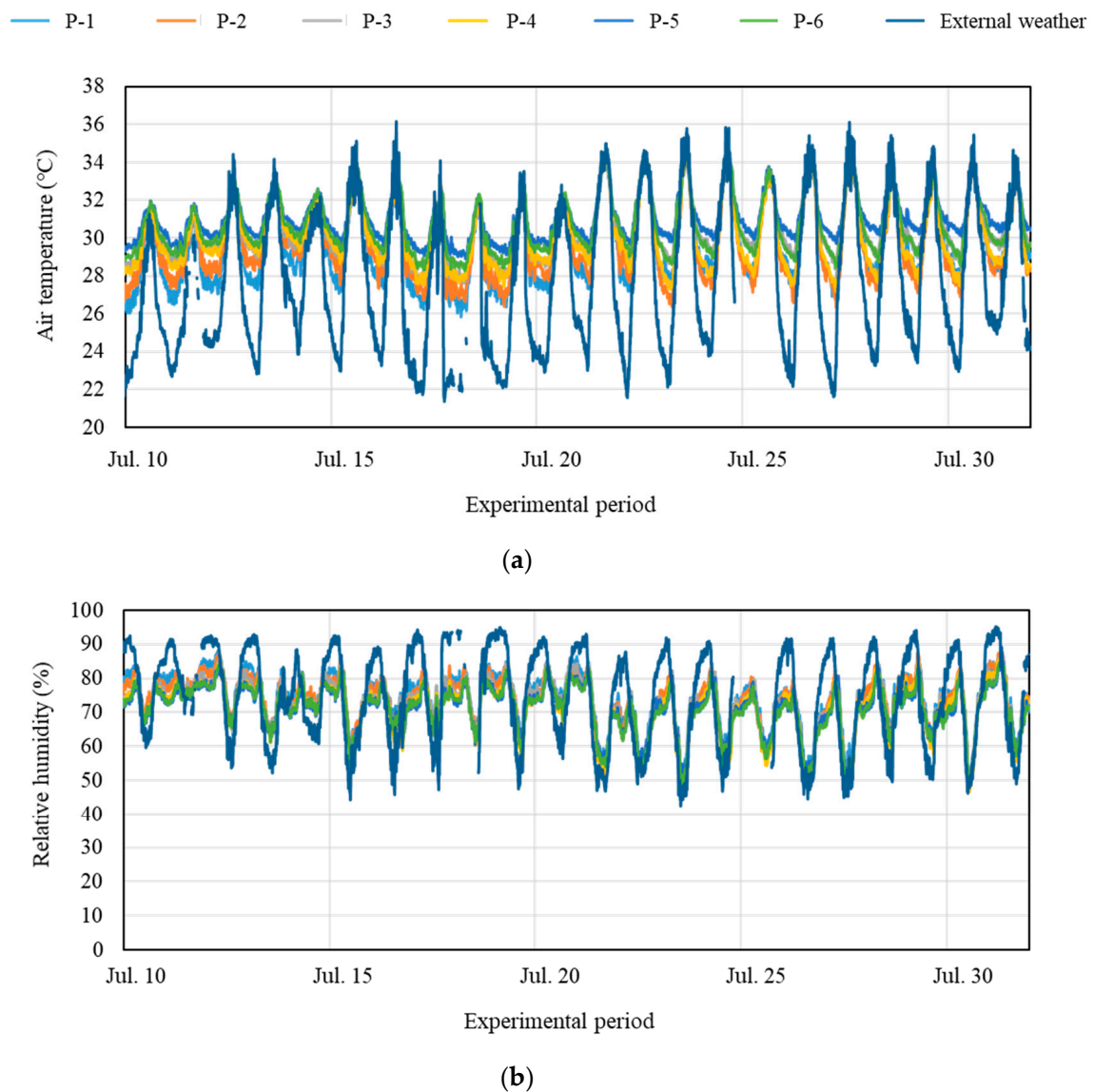


Figure 8. The results of measured air temperature and relative humidity inside the fattening pig house. (a) Air temperature. (b) Relative humidity.

Based on the results, it was expected that the pigs had high-temperature stress. The THI was additionally calculated by simultaneously considering the air temperature and relative humidity, as shown in Figure 9a. The ammonia concentration and ventilation rate inside the fattening pig house were also presented in Figure 9b,c. Accordingly, the measured THI ranged from 78 to 86 during the experimental period. This means that the internal environments were under the alert and danger sections for high-temperature stress on pigs. Although the ventilation rates were maximum in the daytime, the internal air temperature of the fattening pig house was higher than 34 °C. Therefore, it is recommended to install additional cooling systems to relieve the high-temperature stress in summer.

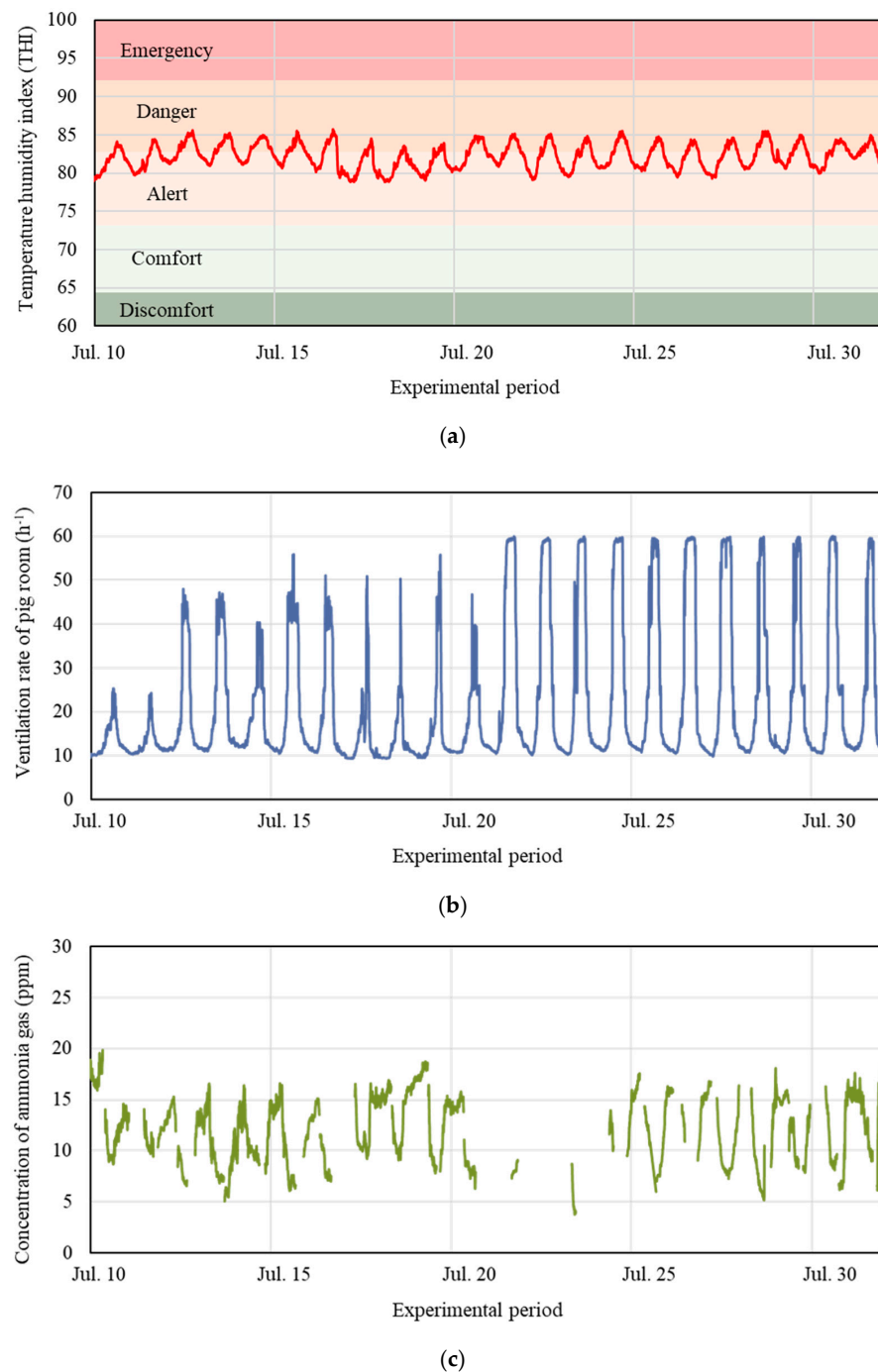


Figure 9. The evaluation results of the breeding environment inside the experimental pig house. (a) Temperature humidity index. (b) Ventilation rate of the pig room. (c) Concentration of ammonia gas inside the pig room.

In the case of ammonia concentration inside the fattening pig house, a relatively high range of 5 to 20 ppm was measured. The ammonia concentration fluctuated with the change in ventilation rates during the day and night. The ammonia concentration increased as the ventilation rate decreased at night. On the other hand, the concentration of ammonia gas was lowered due to the high ventilation rate in the daytime. As a result of the monitoring in the fattening pig house, it was identified that internal environments were poor with a high air temperature, relative humidity, and ammonia concentration. Due to the poor environment inside fattening pig houses, it is hard to accumulate reliable long-term data on the fattening pig house due to sensor failure. Therefore, it is necessary to conduct research for filling in the missing data on the environments inside livestock houses. In this study, the research for the data filling of the missing data as the basic research was conducted using air temperature data, which is one of the most important factors for the environmental control inside fattening pig houses.

3.2. Validation of the Accuracy of the Developed Models

The developed RNN model in this study was validated by comparing air temperature data predicted by the RNN model with the assumed missing data. The air temperature predicted by the RNN model was presented in Figure 10, according to the periodic parameter. The RMSE and MAPE as statistical indices were calculated for the quantitative comparison, as shown in Table 4. When the RNN model was developed by learning only internal air temperature, the value of the RMSE and MAPE were 1.92 °C and 4.70%, respectively. As shown in Figure 10a, the RNN model could not accurately predict the internal air temperature when only air temperature data was considered as the learning data. Therefore, additional processes, such as considering a multi-layer model and adding other parameters, were necessary to improve the accuracy of the RNN model for predicting the missing data of the air temperature inside the fattening pig house. However, the computing loads and learning time could increase for the development of the multi-layer RNN model. Therefore, the RNN model additionally considers the periodic parameter as the learning data. When the RNN model was developed by considering the periodic parameter, the value of the RMSE and MAPE were 1.41 °C and 3.55%, respectively. The accuracy of the RNN model was improved with a 1.15% decrease in the error rate. The tendency of air temperature predicted by the RNN model considering the periodic parameter was fitted with the tendency of air temperature measured in the field experiment. It was suitable to use the developed RNN model for predicting the air temperature inside the fattening pig house within a 3.5% error rate. The developed RNN model in this study could be applied to predict the internal air temperature when the monitoring sensors are not working. Furthermore, the RNN models were expected to be highly applicable because they could be continuously improved the RNN model through learning from the monitoring data.

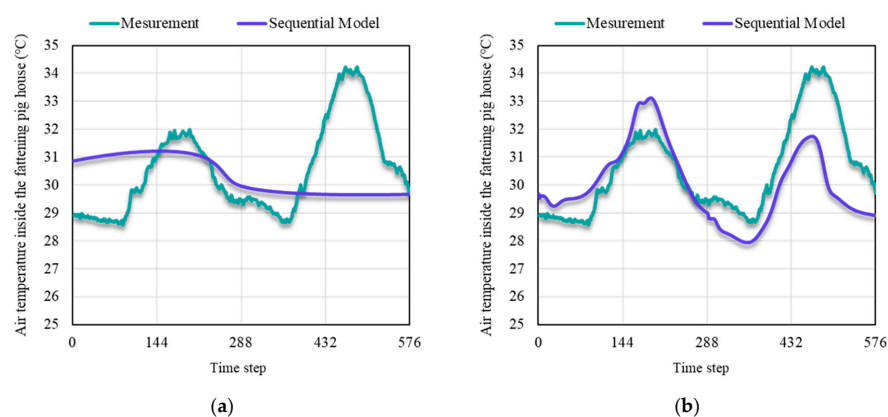


Figure 10. A comparative analysis of the predictive air temperature inside the fattening pig house according to the periodic parameter. (a) Learning without the periodic parameter. (b) Learning with the periodic parameter.

Table 4. Validation of the RNN model for predicting the missing data, according to learning parameters.

Statistical Indices	Learning without the Periodic Parameter	Learning with the Periodic Parameter
RMSE (°C)	1.92	1.41
MAPE (%)	4.70	3.55

3.3. A Comparative Evaluation of the RNN Models according to the Order of Learning Data

The RNN model was developed by learning time series data in sequential and reverse orders. The bidirectional model was additionally developed by combining the sequential model and reverse model to improve the accuracy of the RNN models. The air temperature predicted by the sequential, reverse, and bidirectional models were presented in Figure 11. A comparative analysis was also conducted using the statistical indices of the RMSE and MAPE, according to the order of learning data and the length of the missing data. The increase in the RNN model's accuracy was presented in Table 5, according to the order of learning data.

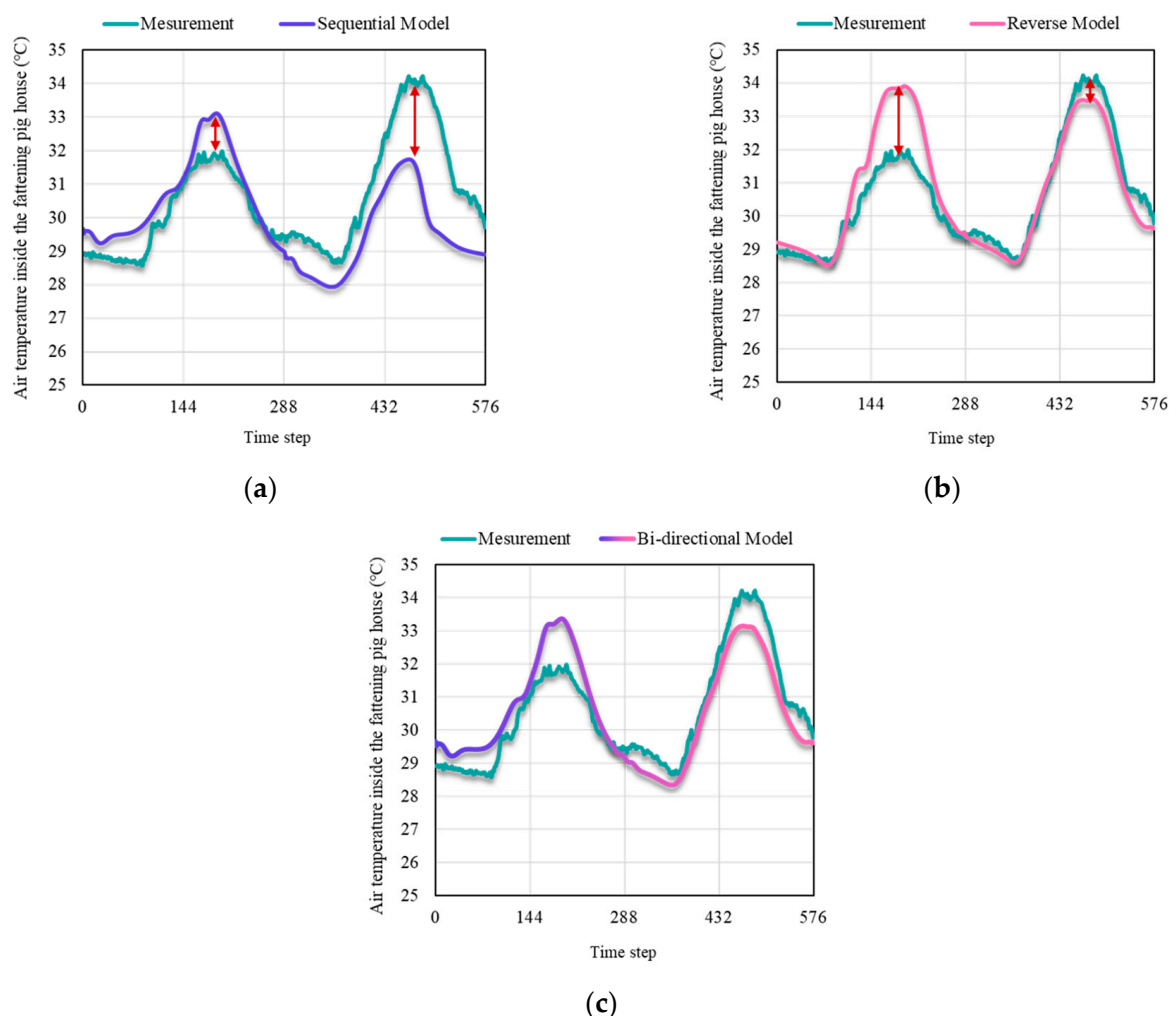


Figure 11. A comparative analysis of the predictive air temperature inside the fattening pig house according to the order of learning data. (a) Sequential model. (b) Reverse model. (c) Bidirectional model.

Table 5. An increase in the RNN model's accuracy according to the order of learning data.

MAPE (%)	The Length of the Missing Data								
	1 hr	2 hr	3 hr	6 hr	9 hr	12 hr	24 hr	36 hr	48 hr
Sequential model (A)	2.15	2.25	2.89	2.94	2.77	2.33	2.14	2.81	3.55
Bidirectional model (B)	0.24	0.77	0.43	0.93	1.06	1.12	1.48	2.60	2.64
Increase in accuracy (A-B)	1.91	1.48	2.46	2.01	1.71	1.21	0.66	0.21	0.91

As shown in Figure 11a, the sequential model was relatively more accurate at the 0–144 time step than at the 144–288 time step. On the contrary, the reverse model was relatively more accurate at the 144–288 time step than at the 0–144 time step, as shown in Figure 11b. As a result of the quantitative comparison (Table 6), the reverse model was more accurate than the sequential model. These results were in agreement with the results of previous studies [39–41]. The sequential model was able to accurately predict the missing data of the internal air temperature with an error rate of 2.14–3.55%, according to the length of the missing data. The reverse model was accurate with an error rate of 0.72–3.11%.

Table 6. The RNN model's accuracy, according to the order of learning data and the length of the missing data.

Sequential order	The Length of the Missing Data								
	1 hr	2 hr	3 hr	6 hr	9 hr	12 hr	24 hr	36 hr	48 hr
RMSE (°C)	0.63	0.67	0.84	0.85	0.82	0.73	0.86	0.87	1.41
MAPE (%)	2.15	2.25	2.89	2.94	2.77	2.33	2.14	2.81	3.55
Reverse order	Sequence length for the LSTM model								
	1 hr	2 hr	3 hr	6 hr	9 hr	12 hr	24 hr	36 hr	48 hr
RMSE (°C)	0.55	0.22	0.65	0.41	0.37	0.72	1.17	1.07	0.83
MAPE (%)	1.88	0.72	2.21	1.27	1.13	1.93	3.11	2.59	1.85
Bidirectional order	Sequence length for the LSTM model								
	1 hr	2 hr	3 hr	6 hr	9 hr	12 hr	24 hr	36 hr	48 hr
RMSE (°C)	0.08	0.25	0.14	0.30	0.35	0.36	0.51	0.87	0.99
MAPE (%)	0.24	0.77	0.43	0.93	1.06	1.12	1.48	2.60	2.64

The accuracy of the bidirectional model was improved and there was a decrease in the error rate of 0.21–2.46% compared with the accuracy of the sequential model. Especially, the shorter the length of the missing data, the higher the accuracy of the bidirectional model. When the length of the missing data was an hour, the values of the RMSE and MAPE for the sequential model were 0.63 °C and 2.15%, respectively. When the length of the missing data was an hour, the values of the RMSE and MAPE for the bidirectional model were 0.08 °C and 0.24%, respectively. When the length of the missing data was 12 h, the values of the RMSE and MAPE for the sequential model were 0.73 °C and 2.33%, respectively. When the length of the missing data was 12 h, the values of the RMSE and MAPE for the bidirectional model were 0.51 °C and 1.48%, respectively. When the length of the missing data was 24 h, the values of the RMSE and MAPE for the sequential model were 1.41 °C and 3.55%, respectively. When the length of the missing data was 24 h, the values of the RMSE and MAPE for the bidirectional model were 0.99 °C and 2.64%, respectively. The accuracies of the bidirectional models with the length of the missing data of 1, 12, and 24 h were improved, as error rates of 1.91, 1.21, and 0.91% were decreased by comparing with the accuracy of the sequential model, respectively.

The accuracy of the bidirectional model was improved by comparing it with the accuracy of the sequential model. Therefore, the missing data could be more accurately predicted using the bidirectional model when the air temperature data are accumulated

after the missing data. Furthermore, the longer the length of the missing data, the higher the applicability of the bidirectional model because the accuracy of the RNN model was lower when the length of the missing data was longer.

3.4. A Comparative Evaluation of the RNN Models according to the Length of the Missing Data

In this study, the accuracies of the developed RNN models were evaluated according to the length of the missing data. Generally, the evaluation of the model should state that a larger number of missing data will result in a lower accuracy of the RNN model. The RMSE and MAPE values of the sequential, reverse, and bidirectional models, according to the length of the missing data, were presented in Figure 12. As the result of the comparative analysis, the tendencies of the sequential and reverse models were not clear, as shown in Figure 12a,b. However, the tendency of the bidirectional model was clear, as shown in Figure 12c. The bidirectional model developed in this study could accurately predict the missing data of air temperature inside the fattening pig house. When the length of the missing data was 6 h, the values of the RMSE and MAPE for the bidirectional model were 0.30 °C and 0.93%, respectively. When the length of the missing data was 24 h, the values of the RMSE and MAPE for the bidirectional model were 0.51 °C and 1.48%, respectively. When the length of the missing data was 48 h, the values of the RMSE and MAPE for the bidirectional model were 0.99 °C and 2.64%, respectively. Therefore, when the length of the missing data was less than 6 h, the bidirectional model could predict the missing data of the internal air temperature within a 1% error rate. When the length of the missing data was less than 24 and 48 h, the bidirectional model could predict the missing data of the internal air temperature within 1.5% and 3% error rates, respectively.

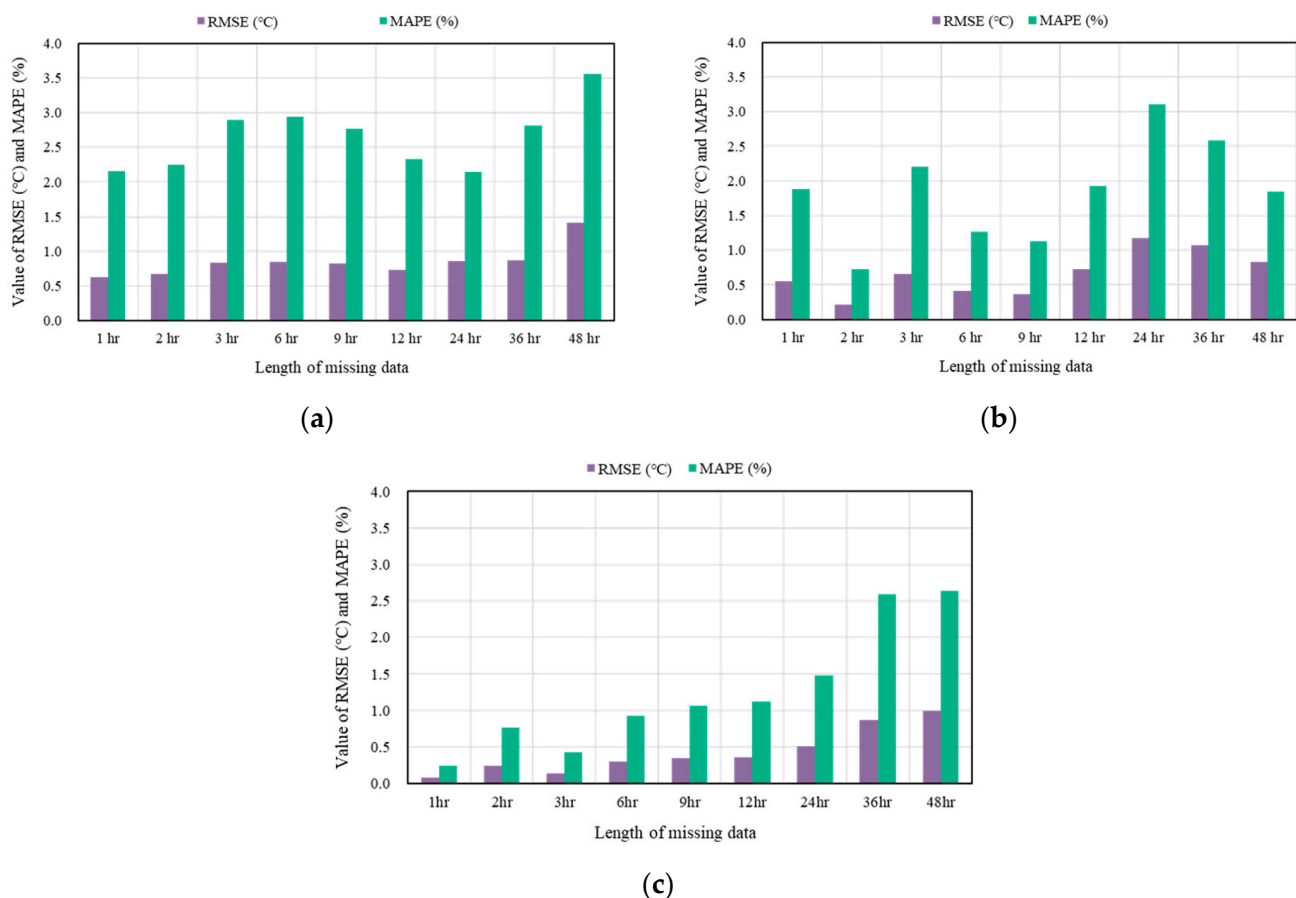


Figure 12. A comparative analysis of the model accuracies according to the length of the missing data. (a) Sequential model. (b) Reverse model. (c) Bidirectional model.

Finally, the RNN model developed in this study could be usefully applied to monitor the internal environments of the fattening pig houses, where it is difficult to accumulate reliable data for the long term due to the poor environments such as high air temperature, relative humidity, and ammonia concentration. RNN models have the advantage that the accuracy of RNN models can be continuously improved by learning from monitoring data. The RNN model could be improved to learn the data of several factors, such as external weather and ventilation rates, which directly impact the internal environments. Since the RNN model developed in this study was developed by learning only temperature data, it is highly applicable to livestock facilities where temperature-based control is used.

4. Conclusions

In this study, the developed RNN models could predict the missing data of air temperature inside the experimental fattening pig house. First, a field experiment was conducted to identify the environmental problems of the fattening pig house. During the field experiment, air temperature, relative humidity, ventilation rate, ammonia concentration inside the fattening pig house, and the external weather, were monitored. From these, RNN models for data filling of the missing data were developed and validated. Based on analyzed monitoring data, it was found that the pigs were expected to experience high-temperature stress. Although the ventilation rates were maximum in the daytime, the air temperature and THI inside the fattening pig house were higher than 34 °C and 85 during the experimental period, respectively. Therefore, it is necessary to conduct research for filling in the missing data on the environments inside livestock houses.

In this study, the research for the data filling of the missing data as the basic research was conducted using air temperature data, which is one of the most important factors for environmental control inside fattening pig houses.

The RNN model for predicting the missing data by learning air temperature data is one of the most important factors for environmental control inside fattening pig houses. Some of the measured air temperature data inside the fattening pig house were assumed to be the missing data. As a result of validating the accuracy of the RNN model considering the periodic parameter, the RNN model for predicting the air temperature inside the fattening pig house was accurate within a 3.5% error rate. Therefore, the RNN model was useful for predicting data to replace missing data when it occurs.

The accuracies of the RNN models were evaluated according to the order of learning data. Based on the result, the accuracy of the bidirectional model was improved by comparing it with the accuracy of the sequential model. Therefore, the missing data could be more accurately predicted using the bidirectional model when air temperature data are accumulated after the missing data. Furthermore, the accuracies of the RNN models were compared and analyzed according to the length of the missing data. The longer the length of the missing data, the lower the accuracy of the RNN model. When the length of the missing data was less than 26, 4, and 48 h, the bidirectional model could predict the missing data of the internal air temperature within 1%, 1.5%, and 3% error rates, respectively.

The main contribution of this study is the development of the RNN model for predicting the time series missing data of the air temperature inside fattening pig houses. The RNN models also have an advantage because the accuracies of the RNN models could be continuously improved by learning the measured data. Furthermore, the RNN models developed in this study have high applicability because the RNN model was developed by learning only air temperature data. In the future, it is possible to improve the accuracy of the RNN model by learning the data of several factors, such as external weather and ventilation rates, which directly impact the internal environments. Research predicting the data of ammonia concentration data is necessary because the ammonia sensor is more vulnerable to poor environments inside fattening pig houses.

Author Contributions: Conceptualization, J.-g.K. and I.-b.L.; data curation, S.-y.L.; methodology, J.-g.K. and S.-y.L.; software, J.-g.K.; supervision, I.-b.L.; visualization, J.-g.K.; writing—original draft, J.-g.K.; writing—review and editing, I.-b.L. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through the Livestock Industrialization Technology Development Program, which was funded by the Ministry of Agriculture, Food, and Rural Affairs (MA-FRA) (321085-5).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All data will be made available upon request to the correspondent author's email with appropriate justification.

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

References

- Ministry of Agriculture Food and Rural Affairs. 2021. Available online: www.mafra.go.kr (accessed on 12 December 2022).
- Korea Labor Institute. *Effects of Smart Farm Activation Policy on Employment*; KLI: Sejong, Republic of Korea, 2019; p. 55.
- Seo, I.-H. Development of Wearable Device for Monitoring Working Environment in Pig House. *J. Korean Soc. Agric. Eng.* **2020**, *62*, 71–81.
- Jinubala, V.; Lawrance, R. Analysis of Missing Data and Imputation on Agriculture Data With Predictive Mean Matching Method. *Int. J. Sci. Appl. Inf. Technol.* **2016**, *5*, 1–4.
- Lokupitiya, R.S.; Lokupitiya, E.; Paustian, K. Comparison of missing value imputation methods for crop yield data. *Env. Off. J. Int. Env. Soc.* **2006**, *17*, 339–349. [[CrossRef](#)]
- Zakaria, W.; Salleh, M.Z. Determination of the best single imputation algorithm for missing rainfall data treatment. *J. Qual. Meas. Anal.* **2016**, *12*, 79–87.
- Staub, B.; Hasler, A.; Noetzli, J.; Delaloye, R. Gap-filling algorithm for ground surface temperature data measured in permafrost and periglacial environments. *Permafr. Periglac. Process.* **2017**, *28*, 275–285. [[CrossRef](#)]
- Rubin, L.H.; Witkiewitz, K.; Andre, J.S.; Reilly, S. Methods for handling missing data in the behavioral neurosciences: Don't throw the baby rat out with the bath water. *J. Undergrad. Neurosci. Educ.* **2007**, *5*, A71.
- Ferrari, G.T.; Ozaki, V. Missing data imputation of climate datasets: Implications to modeling extreme drought events. *Rev. Bras. Meteorol.* **2014**, *29*, 21–28. [[CrossRef](#)]
- Afrifa-Yamoah, E.; Mueller, U.A.; Taylor, S.; Fisher, A. Missing data imputation of high-resolution temporal climate time series data. *Meteorol. Appl.* **2020**, *27*, e1873. [[CrossRef](#)]
- Xie, Q.; Ni, J.-Q.; Li, E.; Bao, J.; Zheng, P. Sequential air pollution emission estimation using a hybrid deep learning model and health-related ventilation control in a pig building. *J. Clean. Prod.* **2022**, *371*, 133714. [[CrossRef](#)]
- Wang, Y.; Zheng, W.; Li, B. A modified discrete grey model with improved prediction performance for indoor air temperatures in laying hen houses. *Biosyst. Eng.* **2022**, *223*, 138–148. [[CrossRef](#)]
- Boomgard-Zagrodnik, J.P.; Brown, D.J. Machine learning imputation of missing Mesonet temperature observations. *Comput. Electron. Agric.* **2022**, *192*, 106580. [[CrossRef](#)]
- Hamzah, F.B.; Mohd Hamzah, F.; Mohd Razali, S.F.; Jaafar, O.; Abdul Jamil, N. Imputation methods for recovering streamflow observation: A methodological review. *Cogent Environ. Sci.* **2020**, *6*, 1745133. [[CrossRef](#)]
- Moon, T.; Lee, J.W.; Son, J.E. Accurate Imputation of Greenhouse Environment Data for Data Integrity Utilizing Two-Dimensional Convolutional Neural Networks. *Sensors* **2021**, *21*, 2187. [[CrossRef](#)]
- Song, W.; Gao, C.; Zhao, Y.; Zhao, Y. A time series data filling method based on LSTM—Taking the stem moisture as an example. *Sensors* **2020**, *20*, 5045. [[CrossRef](#)] [[PubMed](#)]
- Demmers, T.G.; Cao, Y.; Gauss, S.; Lowe, J.C.; Parsons, D.J.; Wathes, C.M. Neural predictive control of broiler chicken and pig growth. *Biosyst. Eng.* **2018**, *173*, 134–142. [[CrossRef](#)]
- Lee, S.-Y.; Lee, I.-B.; Yeo, U.-H.; Kim, J.-G.; Kim, R.-W. Machine Learning Approach to Predict Air Temperature and Relative Humidity inside Mechanically and Naturally Ventilated Duck Houses: Application of Recurrent Neural Network. *Agriculture* **2022**, *12*, 318. [[CrossRef](#)]
- Li, H.; Cryer, S.; Acharya, L.; Raymond, J. Video and image classification using atomisation spray image patterns and deep learning. *Biosyst. Eng.* **2020**, *200*, 13–22. [[CrossRef](#)]
- Liu, D.; Oczak, M.; Maschat, K.; Baumgartner, J.; Pletzer, B.; He, D.; Norton, T. A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. *Biosyst. Eng.* **2020**, *195*, 27–41. [[CrossRef](#)]
- Moon, T.; Choi, H.Y.; Jung, D.H.; Chang, S.H.; Son, J.E. Prediction of CO₂ Concentration via Long Short-Term Memory Using Environmental Factors in Greenhouses. *Hortic. Sci. Technol.* **2020**, *38*, 201–209.
- Wang, L.; Zhang, T.; Wang, X.; Jin, X.; Xu, J.; Yu, J.; Zhang, H.; Zhao, Z. An approach of improved Multivariate Timing-Random Deep Belief Net modelling for algal bloom prediction. *Biosyst. Eng.* **2019**, *177*, 130–138. [[CrossRef](#)]

23. Wu, D.; Wu, Q.; Yin, X.; Jiang, B.; Wang, H.; He, D.; Song, H. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosyst. Eng.* **2020**, *189*, 150–163. [[CrossRef](#)]
24. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [[CrossRef](#)]
25. Pham, V.; Bluche, T.; Kermorvant, C.; Louradour, J. Dropout improves recurrent neural networks for handwriting recognition. In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition, Herssonissos, Greece, 1–4 September 2014; pp. 285–290.
26. Khaki, S.; Wang, L.; Archontoulis, S.V. A cnn-rnn framework for crop yield prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [[CrossRef](#)] [[PubMed](#)]
27. Schmidhuber, J.; Hochreiter, S. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
28. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
29. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
30. Moon, T.; Ahn, T.I.; Son, J.E. Forecasting root-zone electrical conductivity of nutrient solutions in closed-loop soilless cultures via a recurrent neural network using environmental and cultivation information. *Front. Plant Sci.* **2018**, *9*, 859. [[CrossRef](#)]
31. Borchers, M.; Chang, Y.; Proudfoot, K.; Wadsworth, B.; Stone, A.; Bewley, J. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *J. Dairy Sci.* **2017**, *100*, 5664–5674. [[CrossRef](#)]
32. Peng, Y.; Kondo, N.; Fujiura, T.; Suzuki, T.; Yoshioka, H.; Itoyama, E. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. *Comput. Electron. Agric.* **2019**, *157*, 247–253. [[CrossRef](#)]
33. Zhao, K.; Jin, X.; Ji, J.; Wang, J.; Ma, H.; Zhu, X. Individual identification of Holstein dairy cows based on detecting and matching feature points in body images. *Biosyst. Eng.* **2019**, *181*, 128–139. [[CrossRef](#)]
34. Keceli, A.S.; Catal, C.; Kaya, A.; Tekinerdogan, B. Development of a recurrent neural networks-based calving prediction model using activity and behavioral data. *Comput. Electron. Agric.* **2020**, *170*, 105285. [[CrossRef](#)]
35. Peng, Y.; Kondo, N.; Fujiura, T.; Suzuki, T.; Ouma, S.; Yoshioka, H.; Itoyama, E. Dam behavior patterns in Japanese black beef cattle prior to calving: Automated detection using LSTM-RNN. *Comput. Electron. Agric.* **2020**, *169*, 105178. [[CrossRef](#)]
36. Milone, D.H.; Galli, J.R.; Cangiano, C.A.; Rufiner, H.L.; Laca, E.A. Automatic recognition of ingestive sounds of cattle based on hidden Markov models. *Comput. Electron. Agric.* **2012**, *87*, 51–55. [[CrossRef](#)]
37. NRC. *Effect of Environment on Nutrient Requirements of Domestic Animals*; National Research Council: Washington, DC, USA, 1981.
38. St-Pierre, N.; Cobanov, B.; Schnitkey, G. Economic losses from heat stress by US livestock industries. *J. Dairy Sci.* **2003**, *86*, E52–E77. [[CrossRef](#)]
39. Srivastava, N.; Mansimov, E.; Salakhudinov, R. Unsupervised learning of video representations using lstms. In Proceedings of the International conference on machine learning, Lille, France, 6–11 July 2015; pp. 843–852.
40. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 3104–3112.
41. Vinyals, O.; Bengio, S.; Kudlur, M. Order matters: Sequence to sequence for sets. *arXiv* **2015**, arXiv:1511.06391.

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.