Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1605–1614

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL


# EFFECTIVELY TOTAL SUSPENDED PARTICULATE MATTER PREDICTION WITH DEEP LEARNING FOR INTENSIVE WATERFOWL FARMING

XUEHUA ZHAO*, JIAQI DONG†, SHUANGYIN LIU, AND HUILING CHEN

Abstract. Total suspended particulate matter (TSP) is a key parameter of environmental quality assessment in intensive waterfowl farming factories. Too much suspended particulate matter is likely to cause respiratory diseases, reduce immunity, and thus affect the healthy growth of waterfowl. Therefore, accurate prediction of TSP in farming houses is significant for intensive waterfowl farming. In this study, we present an efficient TSP prediction method that fuses wavelet packet decomposition (WPD), gated recurrent unit (GRU), and self-attention mechanism. In the method, the WPD is used to decompose and reconstruct the original TSP data to reduce noise and extract key signals, the GRU is used as the prediction model, and the self-attention mechanism is used to enhance the ability of the GRU. To validate the proposed method, in the experiments, comparisons are made with support vector regression machine, random forest, GRU, WPD-GRU, and EMD-GRU (Empirical mode decomposition, EMD). The experimental results demonstrate that our method has the best prediction accuracy for TSP than other models. The values of evaluation measures are RMSE=0.031, MSE=0.001, $R^2$=0.999, respectively. Our method can provide important decision implications for monitoring TSP in intensive waterfowl farming.

## 1. Introduction

China's waterfowl farming industry has a huge scale in the world, producing the largest number of meat geese and ducks and providing the key opportunity to ease the burden on the food production system and promote the development of the waterfowl economy [4]. With the progress of society, the demand for waterfowl products is increasing, waterfowl farming is changing from traditional family breeding to intensive farming [8]. Intensive waterfowl farming is a complex and dynamic environmental system in which total suspended particulate matter (TSP) is an important factor. The TSP can be generated through feed, feathers, and other substances [7]. When the maximum daily concentration exceeds the environmental quality standards for waterfowl farms, regulatory measures need to be taken to promote the healthy growth of waterfowl. Too much suspended particulate matter

2020 *Mathematics Subject Classification.* 93B45, 93B47.

*Key words and phrases.* Waterfowl farming, total suspended particulate matter, wavelet packet decomposition, self-attention mechanism, gated recurrent unit.

This work is supported by the Ministry of Education Humanities and Social Sciences Research Planning Fund Project (23YJAZH212), Innovation Team Project of Colleges in Guangdong Province (2022KCXTD057), Foundation of State Key Laboratory of Public Big Data (PBD2022-07) and Guangdong Province Vocational College Industry Education Integration Innovation Platform (2024CJPT003).

*†Corresponding authors.

in the house can reduce air quality, some viruses and bacteria that cause respiratory diseases and weakened immunity also rely on suspended particulate matter as a carrier. High concentrations of suspended particulate matter can have serious health effects on livestock and workers [2]. Therefore, timely and accurate prediction of TSP in intensive waterfowl farming farms can effectively prevent diseases and improve the health of waterfowl while protecting the respiratory system of breeders.

In recent years, more and more attention has been paid to the environmental estimation and early warning of intensive farming, and some researchers have conducted lots of research on the environmental impact factors in livestock and poultry houses. Liu et al. used the XGBoost to predict the odor concentration changes in egg houses [6], but the generalization ability of the method is weak. Shen et al. proposed an effective prediction method that fuses empirical modal decomposition (EMD) and neural network for NH3 concentration prediction in pig houses [10], but it is easy to fall into a local minimum. Ortega et al. proposed a temperature prediction method for weaning piglets in seven production cycles [7], however, for non-stationary nonlinear data, the results were not satisfactory. Song et al. proposed an effective method based on kernel principal component analysis and neural networks to predict ammonia concentration changes in cattle barns [12]. Although the aforementioned methods have achieved good results in the prediction of key influencing factors in livestock breeding, there are still problems such as low efficiency, high cost, untimely and instability, so that they are rarely applied in the field of goose breeding.

With the rapid development of deep learning, some neural networks have been proposed for nonlinear prediction [11], such as radial basis networks, deep belief networks, extreme learning machines, and long short-term memory networks (LSTM). Huang et al. used gated recurrent unit neural network (GRU) to predict air pollutant concentrations [3], and experimental results showed that GRU can better capture the dependence of large time steps in time series and reduce the training complexity, but the method lacks attention to the important information of historical data. The attention mechanism can achieve autonomous mining of information while maintaining the integrity of information and enhance key information. Lin et al. showed that LSTMs using attention mechanisms take into account feature correlation and time dependence, so that LSTMs have high accuracy. [5].

During environmental data collection in waterfowl farming houses, noise is generated due to factors such as aging sensors or dirt adhering to the surface. To recover real data and eliminate noise distractions, the researchers have used the EMD, wavelet transform (WT), and independent component analysis to remove noises in recent years, in which the WT is outstanding and has a good noise reduction effect [4]. The WT focuses on arbitrary details of signals and performs multi-resolution time-frequency domain analysis, it is straightforward and fast to compute, but the trial range is very narrow, only for the frequency range of known noise in specific cases. Wavelet packet decomposition (WPD) [1] is a further optimization of wavelet variation and solves the problem of low resolution in the high-frequency region.

Consequently, considering the complementary advantages of GRU and self-attention mechanisms, an efficient TSP prediction method based on WPD, self-attention mechanisms and GRU for intensive waterfowl farming is proposed in this study. In this method WPD first decomposes the signal into wavelet packet coefficients of different frequency bands, and the coefficient matrix is input to GRU with self-attention to predict future changes of TSP. GRU can effectively solve the long-term dependency problem, self-attention can ensure the integrity of information and achieve effectively learn the correlation between data, thereby improving the prediction accuracy of TSP. The proposed method can provide a useful tool for TSP the management and supervision to reduce the risk of waterfowl diseases.

## 2. Model and methods

### 2.1. **Wavelet packet decomposition.**

Wavelet packets are a special linear combination of wavelets, and WPD is a finer signal decomposition method developed on the basis of wavelet decomposition, which overcomes the limitation that wavelet decomposition can only decompose the low-frequency part. WPD can perform good time-frequency localization analysis on signals containing a large amount of high-frequency information, and can effectively solve feature extraction for non-smooth data.

The expression underlying decomposition for original signal is as follows:

$$d_t^{j,2k} = \Sigma p h_{p-2t} d_p^{j+1,2k}, \tag{2.1}$$

$$d_t^{j,2k+1} = \Sigma p g_{p-2t} d_p^{j+1}. \tag{2.2}$$

The basis of wavelet packet reconstruction is as follows:

$$d_t^{j+1,k} = \Sigma p(h_{t-2p} d_p^{j,2k} + g_{t-2p} d_p^{j,2k+1}) \tag{2.3}$$

where $d_t^{j,2k}$ is the input raw signal, $d_p^{j+1,2k}$ and $d_t^{j,2k+1}$ respectively are the decomposed high-frequency and low-frequency signals, where $d$ is the wavelet coefficient, $j$ and $k$ are the numbers of nodes in the single layer of the decomposition, $h$ and $g$ are the filtering generalizations, $p$ and $t$ are the number of high-frequency and low-frequency decomposition layers.

### 2.2. **Gated recurrent unit.**

The GRU is an advanced variant derived from the LSTM and has excellent processing ability for a number of discrete time series. The GRU has only two gates, the reset gate and the update gate. The GRU combines the input and forgetting gates of LSTM into a single update gate, which means that fewer parameters are included and convergence is faster, while maintaining the excellent capabilities of LSTM.

The update gate determines whether information from the previously hidden state is passed to the current hidden state, while the reset gate determines the amount of information that is forgotten from the previously hidden state. The detailed formula for the GRU forward training process is as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \tag{2.4}$$

$$r_t = \sigma(W_r \cdot [h_(t-1), x_t]), \tag{2.5}$$

$$g_t = tanh(W_{gt} \cdot [r_t \bigodot h_(t-1), x_t]), \tag{2.6}$$

$$h_t = (1 - z_t) \bigodot h_{t-1} + z_t \bigodot g_t) \tag{2.7}$$

where $x_t$ and $h_t$ are the input vector and output vector, respectively. $g_t$ is the candidate activation vector, $z_t$ and $r_t$ are the update gate and reset gate, respectively. $\sigma$ is sigmoid activation function, $tanh$ is the hyperbolic tangent activation function of candidate hidden states, $W_z$, $W_r$, $W_{gt}$ are the weight matrices of corresponding states, and $\bigodot$ is two vector dot product.

2.3. **Self-attention mechanism.** Self-attention mechanism is an improvement on the attention mechanism by reducing the reliance on external information and is better at capturing the internal relevance of data. In the attention mechanism, the attention takes place between the query of the source and all elements of the target, whereas, the attention occurs between internal elements of the source in the self-attention mechanism.

The attention mechanism is calculated as follows:

$$Attention(Q, K, V) = softmax\Big(\frac{QK^T}{\sqrt{d_k}}\Big)V \tag{2.8}$$

where $Q \in R^{n*d_k}$, $K \in R^{m*d_k}$, $V \in R^{n*d_v}$, $\sqrt{d_k}$ is the adjustment factor. For self-attention, $Q = K = V$. Self-attention focuses only on the features possessed by input data itself, making it easier to capture inter-dependencies between long distances while increasing computational parallelism.

2.4. **Proposed WPD-self-attention-GRU model for TSP prediction.** To improve TSP prediction accuracy, in this study, we propose an effective TSP prediction method (WPD-self-attention-GRU for short) integrating WPD, self-attention and GRU for intensive waterfowl farming. In WPD-self-attention-GRU, WPD is used to decompose and reconstruct original TSP time series data to reduce the noise and extract key signals. GRU is used as the prediction component which can solve the gradient disappearance and gradient explosion problem. The self-attention mechanism is used to capture the internal correlation of time series data and solve long-term dependence problem. The flow of the algorithm is shown in Figure1.

The data is collected online via the waterfowl farming IoT cloud service platform. However, there are outliers or missing values for TSP time series data due to aging sensors or dirt adhering to the surface, which are repaired and normalized according to Eq.2.9

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2.9}$$

where $x_{max}$ and $x_{min}$ are the maximum and minimum of TSP, respectively.

Then the standardized TSP data is decomposed into wavelet packets of three layers with a total of eight components having different feature scales by WPD since a wavelet packet with three decomposition layers can approximate any nonlinear function.

The decomposed signals are reconstructed singly to form new low-frequency sequences that can describe trend changes of TSP sequences, and high-frequency sequences that retain different information. The reconstructed data are divided into eight different scales.

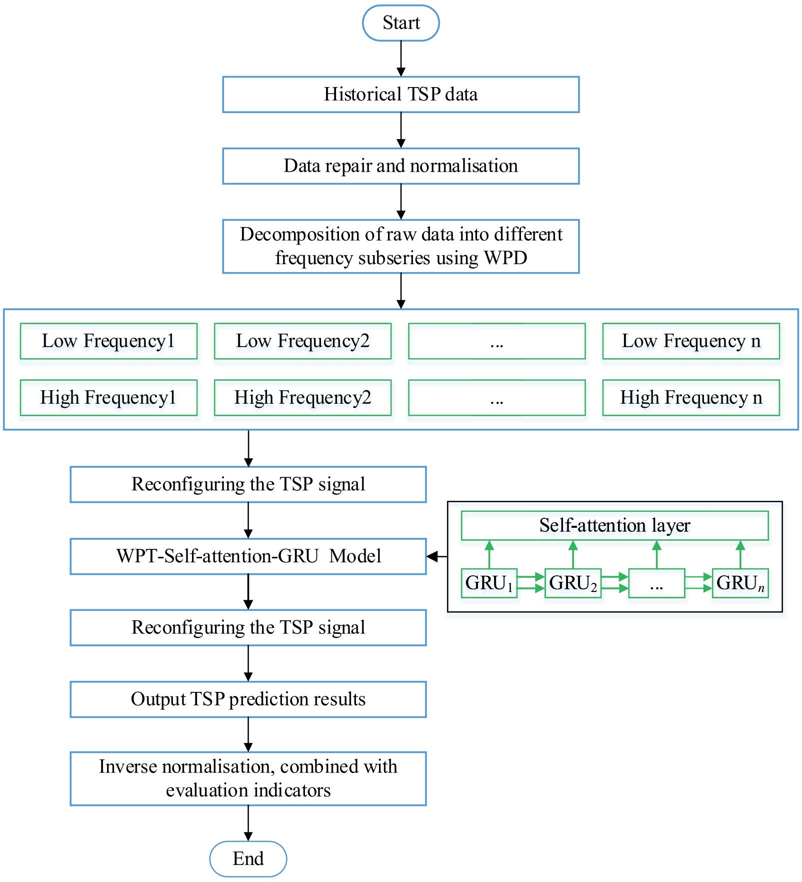


FIGURE 1. Flow chart of WPD-self-attention-GRU algorithm

The prediction target is the TSP value for the next 20 minutes. The input of the model is historical TSP data and the output is the predicted value of TSP. Based on the superiority of GRU in time series prediction, the GRU is used to construct a TSP-based time series prediction model. The introduction of self-attention in the GRU facilitates capturing time-dependent features. The predicted value is back-normalized to obtain the final values of TSP.

## 3. EXPERIMENTS AND APPLICATIONS

3.1. **Data sources and data collection.** Environmental data is collected from the particular waterfowl farming base in Haifeng County, Shanwei City, China. 1000 lion-headed geese are raised on this base. A semi-enclosed waterfowl house with 25m in length and 16m in width in the base is selected as the source of data collection in our experiments. The base includes five parts: geese house, egg laying area, isolation area, exercise area and bathing pool. The plan view of the IoT-based waterfowl farming base is shown in Figure 2. The walls of the geese house are constructed of rock wool sandwich-colored steel panels with built-in fans, fill lights and various sensors. The geese house is mainly naturally ventilated, supplemented

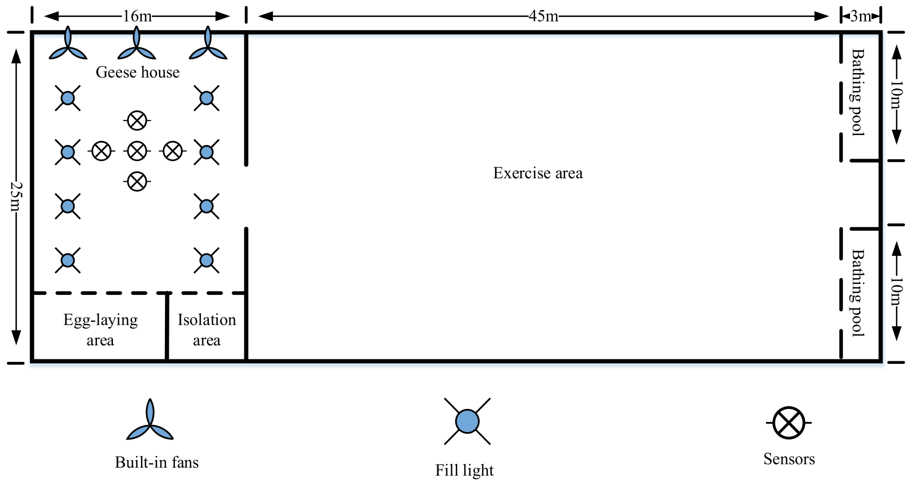


FIGURE 2. The view of the IoT-based waterfowl farming base

by mechanical ventilation. Natural and supplemental lights are used to provide light for lighting. Fans and spray cooling equipment are used to cool the house. A variety of sensors are deployed inside the geese house to monitor the environmental parameters of the geese house online. The geese house is built on higher ground to prevent rainwater from flowing into the house.

The TSP in geese house is collected every 20 minutes via the IoT cloud service platform. The collection period is from 1 July to 28 September 2023, the collected data included 6,480 items. In our experiments, the data is divided into training set and test set. The first 80% of data, 5180, is used as training set, the rest 20%, 1296, is used as test set.

3.2. **Parameter setting.** All the experiments are run on a computer with an Intel I5-5200U processor and an operating system of Windows 10. The Anaconda3 is chosen as the interpreter and Python 3.7 is the programming language, and Keras is used to train the model.

After repeated experiments, the optimal parameters of GRU are selected as follows: the number of neurons in the hidden layer is 64, the number of layers is 2, the learning rate is 0.0001, the batch size is 64, the number of iterations is 100, the activation function is Relu, the optimization method is Adam, the loss function is mean squared error (MSE), the dropout is 0.2, and the self-attention layer activation function is sigmoid.

In our experiments, root mean square error (RMSE), MSE and coefficient of determination ($R^2$) are used to evaluate the performance of methods. The smaller values of RMSE and MSE indicate a smaller deviation between actual and predicted values of the model. For $R^2$, close to 1 indicates a strong correlation between predicted and actual values. The SVM, RF, GRU, EMD-GRU and WPD-GRU are selected for comparisons.

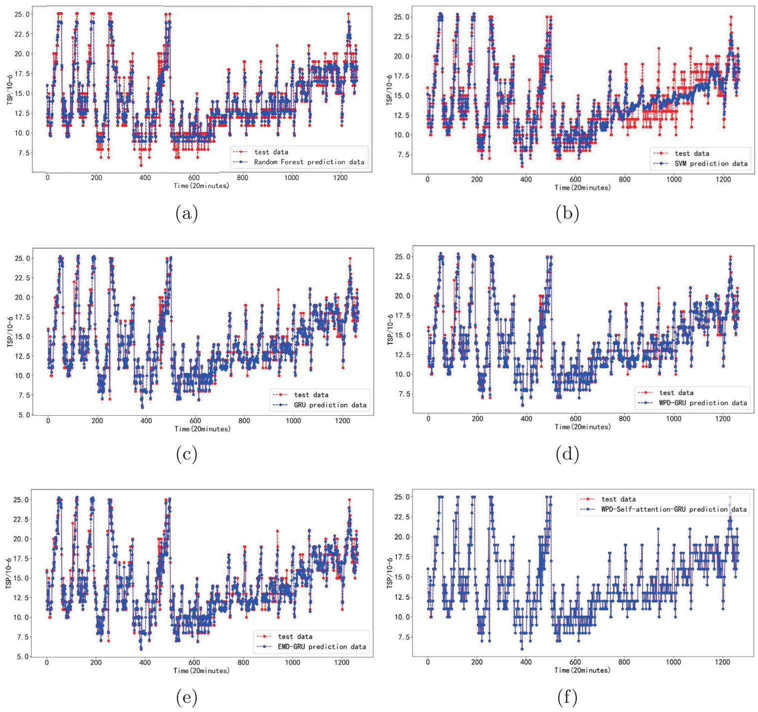


FIGURE 3. Comparisons on different models

TABLE 1. Evaluation results of 6 models

| Model | RMSE | MSE | $R^2$ |
|---|---|---|---|
| RF | 0.364 | 0.133 | 0.865 |
| SVM | 0.26 | 0.068 | 0.931 |
| GRU | 0.152 | 0.023 | 0.976 |
| EMD-GRU | 0.147 | 0.022 | 0.978 |
| WPD-GRU | 0.122 | 0.015 | 0.985 |
| WPD-self-attention-GRU | 0.031 | 0.001 | 0.999 |

## 4. Results and discussion

The experimental results are shown in Figure 3. As shown in Figure 3(a), (b) and (c), traditional machine learning methods have some disadvantages of low prediction accuracy and poor robustness, while neural networks overcome these shortcomings. As shown in Figure 3(c), (d) and (f), the prediction accuracy of the GRU after noise reduction is higher. The reason is that TSP is susceptible to interference from external factors such as feed, excrement, and feathers, and is nonlinear and unstable. WPD and EMD can reduce noise interference and extract useful information, but the noise reduction effect of EMD is not as good as that of WPD, because EMD is easy to cause modal confusion. As can be seen in Figure 3(e) and (f), the proposed method is the best because self-attention can capture the internal correlation of time series data.

TABLE 2. Evaluation results of 6 models

| Model | RMSE | MSE | $R^2$ |
|---|---|---|---|
| RF | 0.4366 | 0.1906 | 0.8056 |
| SVM | 0.4136 | 0.1711 | 0.8245 |
| GRU | 0.3962 | 0.1570 | 0.8425 |
| EMD-GRU | 0.3978 | 0.1583 | 0.8415 |
| WPD-GRU | 0.3925 | 0.1541 | 0.8455 |
| WPD-self-attention-GRU | 0.1816 | 0.0330 | 0.9664 |

The values of RMSR, MSE and $R^2$ are listed in Table 1 for six prediction methods. For the proposed method, RMSE, MSE and $R^2$ are 0.031, 0.001 and 0.999, respectively. As we can see in Table 1, the proposed method has the best values of evaluation. For single models RF, SVM and GRU, traditional machine learning models have low prediction accuracy and some lag. The GRU has better overall performance, the values of RMSE and MSE are reduced by 58.24% and 82.71% compared to the RF, and by 41.54% and 66.18% compared to the SVM. this shows that the GRU can better explore the nonlinear variation pattern of TSP. Compared with the standard GRU, WPD-GRU reduced RMSE and MSE by 19.74% and 34.78%, respectively. The WPD-GRU showed 17.01% and 31.82% reduction in RMSE and MSE, respectively, compared to EMD-GRU. The reason is that the WPD has excellent essential feature extraction capability to reduce the negative impact of noise. After adding the self-attention mechanism, its error is close to 0. Compared with WPD-GRU, RMSE and MSE are reduced by 74.59% and 93.33%, respectively, and improved by 1.42%. The reason is that self-attention enables the model to focus on key information in the feature sequence and make full use of the location information of time series.

For further evaluating the generalization ability of the methods, we test them on the rebuilt dataset according to a 60-minute interval. For the rebuilt dataset, the training set contains 1726 items and the test set contains 432 items. The values of RMSR, MSE and $R^2$ are listed in Table 2. We can see that our method has the best performance among six methods. It indicates that the proposed method has the excellent generalization ability.

Consequently, the prediction accuracy and generalization ability of WPD-self-attention-GRU are higher than those of other methods, and the trend of TSP in waterfowl houses can be accurately grasped.

## 5. CONCLUSIONS

TSP is one key environmental factor affecting the growth of waterfowl, and high TSP concentration easily leads to the decreased immunity and respiratory diseases. To prevent the waterfowl diseases, the WPD-self-attention-GRU is proposed to predict TSP. In the proposed method, WPD can separate the trend component and noise component of original TSP data, and reduce the adverse effect of noise signals on the prediction results. The self-attention can directly capture global connections and improve the training efficiency of the model. Self-attention-GRU can adequately extract features and solve long-term dependency problems. The prediction results of the proposed method are closer to true values than SVM, RF, GRU, EMD-GRU and WPD-GRU. The proposed method has high prediction accuracy and strong generalization ability. It can provide a reliable reference for the optimal control strategy of TSP in intensive waterfowl farming.

## References

[1] M. Ahmadipour, M. M. Othman, M. Alrifaey, R. Bo, and C. K. Ang, *Classification of faults in grid-connected photovoltaic system based onwavelet packet transform and an equilibrium optimization algorithmextreme learning machine*, Measurement **197** (2022): 111338.

[2] A. P. Fernandez, T. G. Demmers, Q. Tong, A. Youssef, T. Norton, E. Vranken, and D. Berckmans, *Real-time modelling of indoor particulate matter concentration in poultry houses using broiler activity andventilation rate*, Biosystems Engineering **187** (2019), 214–225.

[3] G. Huang, X. Li, B. Zhang and J. Ren, *Pm2. 5 concentration forecasting at surface monitoring sites using gru neural network based onempirical mode decomposition*, Science of the Total Environment **768** (2021): 144516.

[4] J. Huang, S. Liu, S. G. Hassan and L. Xu, *Pollution index of waterfowlfarm assessment and prediction based on temporal convoluted network*, Plos one **16** (2021): e0254179.

[5] J. Lin, J. Ma, J. Zhu and Y. Cui, *Short-term load forecasting based on lstm networks considering attention mechanism*, International Journalof Electrical Power and Energy Systems **137** (2022): 107818.

[6] Y. Liu, Y. Zhuang, B. Ji, G. Zhang, L. Rong, G. Teng and C. Wang, *Prediction of laying hen house odor concentrations using machinelearning models based on small sample data*, Computers and Electronicsin Agriculture **195** (2022): 106849.

[7] J. A. Ortega, E. Losada, R. Besteiro, T. Arango, M. J. Ginzo-Villamayor, R. Velo, M. D. Fernandez and M. R. Rodriguez, *Validation of an autoregressive integrated moving average model for the prediction of animal zone temperature in a weaned piglet building*, Biosystems Engineering **174** (2018), 231–238.

[8] A. Pal, A. P. Jackson, A. Urrutia, K. S. Macklin, S. B. Price, R. J. Buhr and D. V. Bourassa, *Research note: Bacterial composition ofsettled dust during growout of broiler chickens*, Poultry Science **101** (2022): 101602.

[9] Y. Qian, K. Song, T. Hu and T. Ying, *Environmental status of livestock and poultry sectors in china under current transformation stage*, Science of the Total Environment **622** (2018), 702–709.

[10] W. Shen, X. Fu, R. Wang, Y. Yin, Y. Zhang, U. Singh, B. Lkhagva and 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** (2019), 297–305.

[11] L. Shi, H. Zhang, X. Xu, M. Han and P. Zuo, *A balanced social lstm for pm2. 5 concentration prediction based on local spatiotemporalcorrelation*, Chemosphere **291** (2022): 133124.

[12] L. Song, Y. Wang, B. Zhao, Y. Liu, L. Mei, J. Luo, Z. Zuo, J. Yi and X. Guo, *Research on prediction of ammonia concentration in qpsorbf cattle house based on kpca nuclear principal component analysis*, Procedia Computer Science **188** (2021), 103–113.

*Manuscript received February 10, 2024*
*revised September 19, 2024*

X. Zhao
Shenzhen Institute of Information Technology, Shenzhen 518172, China; State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China
*E-mail address*: lcrlc@sina.com

J. Dong
School of Modern Information Industry, Guangzhou College of Commerce, Guangzhou 511363, China
*E-mail address*: 1786975126@qq.com

S. Liu
College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
*E-mail address*: hdlsyxlq@126.com

H. Chen
Department of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, 325035, China
*E-mail address*: chenhuiling.jlu@gmail.com