Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

journal homepage: www.keaipublishing.com/en/journals/information-processing-in-agriculture/

A state of art review on time series forecasting with machine learning for environmental parameters in agricultural greenhouses

Gedi Liu [a,b,c,d,1], *Keyang Zhong* [a,b,c,d,1], *Huilin Li* [a,b,c,d], *Tao Chen* [e], *Yang Wang* [a,b,c,d,*]

[a] *National Innovation Center for Digital Fishery, China Agricultural University, 100083 Beijing, China*
[b] *Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, China Agricultural University, 100083 Beijing, China*
[c] *Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, China Agricultural University, 100083 Beijing, China*
[d] *College of Information and Electrical Engineering, China Agricultural University, 100083 Beijing, China*
[e] *Department of Chemical and Process Engineering, University of Surrey, GU2 7XH Guildford, United Kingdom*

A R T I C L E I N F O

Article history:
Received 3 May 2022
Received in revised form
19 October 2022
Accepted 21 October 2022
Available online xxxx

Keywords:
Horticultural greenhouse environment
Time series algorithms
Prediction
Deep neural networks

A B S T R A C T

Agricultural greenhouse production has to require a stable and acceptable environment, it is therefore essential for future greenhouse production to obtain full and precisely internal dynamic environment parameters. Dynamic modeling based on machine learning methods, e.g., intelligent time series prediction modeling, is a popular and suitable way to solve the above issue. In this article, a systematic literature review on applying advanced time series models has been systematically conducted via a detailed analysis and evaluation of 61 pieces selected from 221 articles. The historical process of time series model application from the use of data and information strategies was first discussed. Subsequently, the accuracy and generalization of the model from the selection of model parameters and time steps, providing a new perspective for model development in this field, were compared and analyzed. Finally, the systematic review results demonstrate that, compared with traditional models, deep neural networks could increase data structure mining capabilities and overall information simulation capabilities through innovative and effective structures, thereby it could also broaden the selection range of environmental parameters for

* Corresponding author at: College of Information and Electrical Engineering, China Agricultural University, 100083 Beijing, China.
E-mail addresses: andy_yangwang@cau.edu.cn, wanghongyang1767@gmail.com (Y. Wang).
[1] G.D. Liu and K.Y. Zhong contributed equally to this work.
Peer review under responsibility of Peer review under the responsibility of China Agricultural University.

Abbreviations: SLR, Systematic literature review; SOTA, State of the art; EC, Exclusion criteria; RQ, Research question; CFD, Computational fluid dynamics; DNN, Deep neural network; ARX, Autoregressive models with external input; ARMAX, Autoregressive moving average models with external input; IARX, Incremental autoregressive models with external variables; ARIMA, Autoregressive integrated moving average model; GA, Genetic algorithm; BP, Back Propagation; SVM, Support vector machines; SVMR, Support Vector Machine Regression; RCB, Randomized Complete Block; LSSVM, Least square support vector machines; GSA, Gravitational search algorithm; MLP, Multilayer perceptron; PCA, Principal Component Analysis; ELM, Extreme learning machine; R-ELM, Extreme learning machine with radial basis activation function; BNN, Bayesian Neural Network; RBNN, Radial Basis Neural Network; RNN, Recurrent neural network; TCN, Temporal convolutional network; LI, Linear interpolation; OS, Random oversampling; DSPT, Dual-stage two-phase; CNN, Convolutional neural network; LSTM, Long- and short-term memory neural network; GRU, Gated recurrent unit; GNN, Graph neural network; MAE, Mean Absolute Error; MSE, Mean Square Error; AME, Average Magnitude of Error; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; R^2, Determination Coefficient; MA, Moving averages

https://doi.org/10.1016/j.inpa.2022.10.005

agricultural facilities and achieve environmental prediction end-to-end optimization via intelligent time series model based on deep neural networks.

Contents

1. Introduction

Recently, the frequent occurrence of extreme weather caused by climate change has seriously affected agricultural production [1]. Meanwhile, the global population will reach about 9.7 billion in 2050 [2]. It means the demand for food continues snowballing, and agriculture is now under increasing pressure [3]. Improving agriculture production efficiency and crop yields by adopting precision or intelligent farming is a wise response.

Facility agriculture has developed rapidly nowadays, as lots of researchers have used advanced computer and automation technologies [4]. Compared with traditional field agriculture, facility agriculture has more tremendous advantages. Through computer auto-control technology, facility agriculture can calculate optimum conditions for agricultural conditions and improve agricultural yield [5]. Facility agriculture provides suitable growth conditions for crops throughout the year by stimulating and controlling the growing environment of crops [6,7]. In addition, facility farming can overcome the regional limitations of field farming and provide employment opportunities [8]. Because of the sensitivity of agricultural production to the environment, the intelligent prediction model was introduced, formulated coping strategies, and provided warning and guidance information for growers [9]. The development of intelligent models promotes the continuous improvement of prediction accuracy. At the same time, the diversified application of models can promote the development of intelligent control algorithms and provide favorable conditions for realizing the automation of algorithms and accurate control of the agricultural environment [10]. Therefore, it is necessary to develop an intelligent environmental prediction model for agricultural facilities.

In recent decades, environmental prediction models of facility agriculture have been studied mainly based on physical processes, and quantitative analysis has been carried out based on the conservation of energy and mass [11]. The climate affects crop production, and researchers have combined climate factors with physical models [12]. [13] uses the physical method of computational fluid dynamics (CFD) to make up for the deficiency of modeling the spatial distribution of environmental parameters. The result shows that it is a powerful simulation tool for agricultural facility environments. However, these physical process-centered prediction models focus on the self-nature of greenhouse structure and materials and lack consideration of diversified environmental parameters and crop factors, making it difficult to establish accurate environmental prediction models. The environmental parameters of agricultural facilities, such as temperature and relative humidity, are characterized by periodicity and fluctuation [14], which needs time series prediction models to fit dynamically. Traditional time series prediction models can be cyclical and simple in fitting time-series trends. Based on Machine Learning and Deep Learning, time series prediction models can be linear and nonlinear function approximations for cyclical volatility and trend sequence [14,15]. Recently, time series models have evolved in depth. The Deep Learning models can search for the internal relationship between input parameters through massive calculation. The RNN and LSTM models can explore the time dependence relationship, the Attention mechanisms, and Graph neural network models respectively focus on the integrity of time and the hidden data structure relationships between parameters [16,17].

At present, researchers have investigated environmental prediction models for agricultural facilities and wrote related review papers. Lopez-Cruz et al. reviewed the greenhouse climate dynamics model based on physical processes [19] without considering the complex relationship of parameters. Deb et al. overviewed energy consumption prediction through Machine learning simulation of the built environment, focusing on single time series data analysis [20]. Compared with our study, this review only discusses the time series algorithm based on machine learning. Additionally, there is no further analysis of the application of more advanced algorithms such as deep learning, and the impact of external variables on the environment prediction of agricultural facilities is rarely considered. Chlingaryan and van Klompenburg summarized the crop yield prediction based on the Machine learning time series model [21,22]. Although they reviewed the application of the deep neural network model, they did not summarize the development of the time series algorithm from the nature of data structure. Guo et al. summarized the modeling of facility agricultural environment from three perspectives: Mechanism, time series, and Machine Learning methods [1]. Although this review elaborated on applying the time series model to the development of the agricultural facility environment, the advantages and disadvantages of each model have not been analyzed. Zhang et al. reviewed the methods of energy-saving control strategies in facility agriculture [23] without summarizing the application of the time series model to energy-saving strategies for facility agriculture. In summary, there are still the following limitations in environmental prediction by time series models of agricultural facilities:

- The application of time series models to predicting the environment of facility agriculture is rarely reviewed.
- The environmental prediction model of agricultural greenhouse is almost not linked to the data characteristics of environmental parameters.
- It lacks a reasonable idea to summarize the development path of the agricultural facility environment prediction model.
- The current summary of time prediction models in facility agriculture lacks advanced and forward-looking.

To outline the application of the time series prediction model in the agricultural facility, we conducted a systematic literature review (SLR). Following the SLR approach, we established a review protocol by specifying the research questions, collecting the appropriate research methods from the applicable database, showing the research scope of multiple screening indicators, and conducting the quantitative analysis. Through planning a systematic literature review, this research work stages in SLR were clearly explained, and the research results were transparent and repeatable [18]. The purpose of this review is to summarize the development environment prediction models for agriculture facilities and the evolutionary development of time series models. Compared with the above review work, this review has the advantage of comprehensively analyzing the development history of intelligent models from the use of data and information strategies, carrying out accuracy generalization evaluation, and then discussing the influence of parameter and time step selection on models. Through systematic induction and study of deep neural networks, the effectiveness of the new model is summarized, and the future application of the model is forecasted.

Section 2 discusses the methodology of this review. Section 3 presents the SLR's valuable results and analyzes the model application in the facility agricultural environment. A discussion on this review has been demonstrated in Section 4, and Section 5 concludes this work.

2. Methodology

2.1. Review protocol

According to a review guideline titled "Guidelines for performing Systematic Literature Reviews in Software Engineering" provided by Kitchenham, redacting review covers three steps, as shown in Fig. 1[24]. First, determine the need for review writing and observe no review related to the time series prediction of greenhouse environmental parameters. The research questions, related keywords, and publication databases were determined. When conducting the review, all databases were consulted to select relevant studies. Selection criteria for primary studies and data synthesis were then determined. We implemented the specified dissemination mechanism in the reporting phase, formatted the main report, and evaluated the information.

2.2. Research questions

This study aims to gain an in-depth understanding of the research on environmental parameter prediction of agricultural facilities and analyzes it from four research questions defined below.

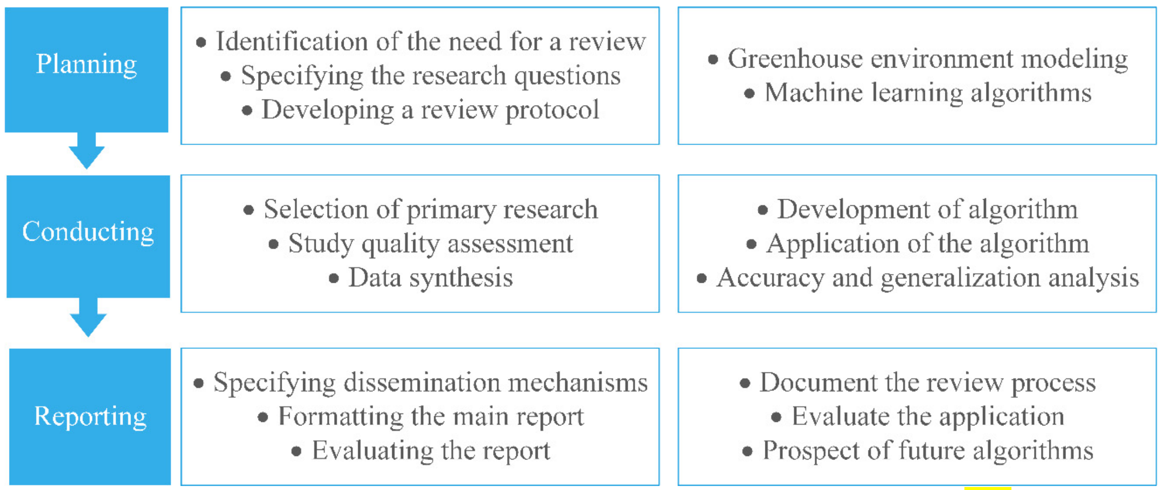

Fig. 1 – Steps involved in planning the systematic literature review (SLR) [24].

- RQ1- Which parameters are paid attention to by researchers in studying time series prediction of environmental parameters of agricultural facilities?
- RQ2- How to evaluate the rationality or accuracy of the time series environmental parameter model developed in agricultural facilities?
- RQ3- How to use models to efficiently utilize data to predict the time series of the environmental parameters in agricultural facilities?
- RQ4- What are the latest models or frontier fields in time series prediction of environmental parameters of agricultural facilities?

2.3. *Search strategy*

A systematic approach reduced the volume of literature to papers relevant to the SLR scope. In order to enhance the standardization of the search process, we simplified and unified the search string to guide subsequent researchers. This final search string is as follows: ((“Machine Learning” OR “Deep Learning” OR “Autoregression”) AND (“Time series” OR “Prediction”) AND (“Agricultural facilities” OR “Greenhouse” Not “gas”)). After executing this search string, 286 studies were retrieved. The search string was used to search in the abstract, title, and keywords fields.

2.4. *Selection criteria*

The literature was analyzed against exclusion criteria to exclude studies irrelevant to the established research topic. The exclusion criteria (EC) are as follows:

EC.1: Publication is not related to the prediction of environmental parameters in agricultural facilities.
EC.2: Publication is duplicated or retrieved from another database.
EC.3: Publication is a survey or review paper.
EC.4: Full text of the study is not available.
EC.5: Publication is not peer-reviewed.

After all exclusion criteria were applied, 61 articles were left worth studying. As illustrated in Table 1, most papers were retrieved from the Google Scholar and Web of Science databases. Table 2 demonstrates the distribution of publications including journals and conference proceedings. During data synthesis, valid information extracted from the article was tabulated in a manner consistent with the statistical information in literature, as shown in Appendix.

3. Results

System planning was carried out according to the review protocol provided by [24]. The detailed search results are stated in Appendix, which indicates the publication year, title, and algorithms used in these publications.

To address the first research question (RQ1), sort and rearrange the selected articles, and perform statistics on the indoor and outdoor parameters used by the model.

To address the second research question (RQ2), the accuracy and generalization of the model are evaluated based on the model classification of the effectiveness of information utilization. Each article has different evaluation methods for the model, and most of them are MAE, MSE, and R^2. All the models have high accuracy, and the accuracy of the new models has been improved. According to statistics, 51 % of models

Table 1 – Database distribution of publications.

Database	Number
Google Scholar	21
IEEE	5
Science direct	6
Web of science	23
Others	6

Table 2 – Distribution of publications.

Publications	Proportion
Applied Energy	10 %
Building and Environment	10 %
Computers and Electronics in Agriculture	25 %
Information Processing in Agriculture	10 %
Sensors	15 %
Transactions of the Chinese Society of Agricultural Engineering	15 %
Transactions of the Chinese Society for Agricultural Machinery	10 %
Others	5 %

lack the generalization of multiple environments, and most models are designed for experimental stations or single laboratories, which lack generalization at the application level of agricultural facility forecasting. 21 % of the models initially had the ability to generalize multiple scenarios.

To address the third research question (RQ3), re-summarize each model from the effectiveness of information utilization, as shown in Fig. 2. The general rule is that the model's efficiency roughly follows the year it was developed and is accompanied by the development of computer computing power. The effectiveness of the model's use of data and the depth of exploration of the data structure continues to improve, and researchers are urged to improve their data collection capabilities.

To address research question four (RQ4), focus on applying Deep learning models in facility agricultural environment modeling. The forefront of current applications lies in the Attention mechanism and Graph neural network. It starts from the model principle and application level, studies advanced models, and anticipate future applications.

3.1. Explore the self-variation of indicators

The classical time series prediction model focuses on the self-changing trend of the detection index. Due to its explainability, it is often used to predict environmental parameters in agricultural facilities. An experiment comparing ARX with ARMAX was done by [25]. The results showed that temperature was the most influential parameter, and ARX performed better than ARMAX. [26] compared the performance of IARX models and ARX models. The IARX model has fewer coefficients, a simple structure, less computation, and accurate results. [27] compared ARX and ARMAX. This experiment provided an experience for the prediction of a controllable environment. [28] concluded that the ARX model based on the GAs algorithm can better fit the air temperature change in the greenhouse. Some scholars combine the method with a neural network [29], combining the ARMAX model with the neural network into NNARMAX, which has good versatility and adaptability [10]. Unlike the single model used before, a temperature prediction model was proposed based on the ma-Arima-GASVR combination method. Experimental results show that the combination method is more efficacious [30].

3.2. For exogenous factors influence

The limitation of autoregression is that it can only simulate linear functions or periodic laws. Machine learning can intelligently explore the relationship between parameters. Its internal principles and application effects can be divided into three application stages: primary nonlinear fitting, multi-dimensional physical division, and artificial intelligence network modeling.

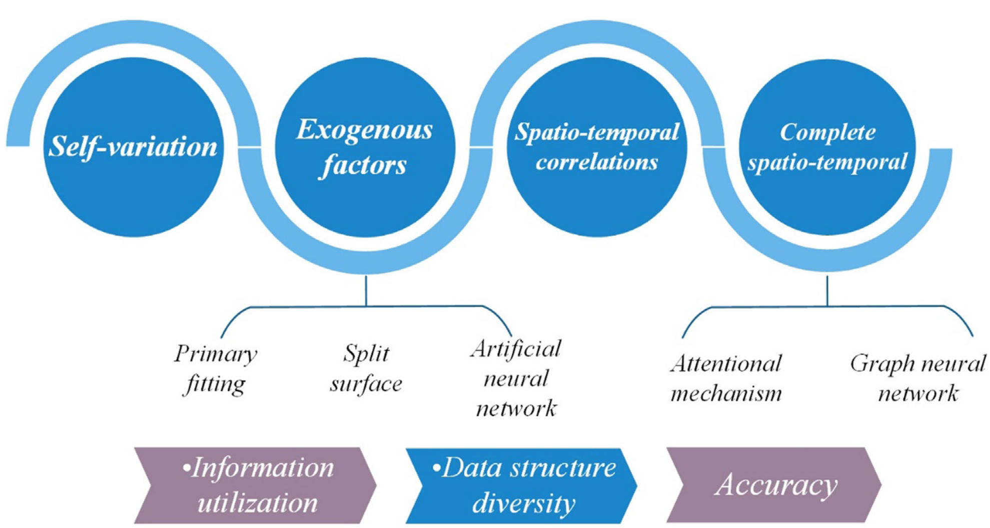

Fig. 2 – Development history of information utilization perspective model.

3.2.1. *Consideration of external factors in the primary stage*

In the initial phase, researchers use a variety of nonlinear functions to fit the data. There are multivariable regression and Fourier function regression in the prediction of environmental parameters of agricultural facilities.

Multivariate Regression: Multivariate regression establishes variability equations to search for influencing parameters in agricultural facilities. [31] use Multivariate Regression to explore the effects of feed consumption and energy consumption, [32] uses Multivariate Regression to evaluate the heat demand.

Fourier function regression: The Sine function was used to improve Fourier function regression. [32] used 12 Fourier series and proposed a least-squares extended Fourier series model combining the least-squares coefficient method.

3.2.2. *Automatic exploration of factor relationships*

In the intermediate stage, support vector machines (SVM) or dimensionality reduction algorithms map the parameter related to a high-dimensional space and fit the complex greenhouse environment by finding the best separation surface in the multi-dimensional space.

Support Vector Machine Regression (SVMR): SVMR minimizes the loss function to maximize the distance between the farthest points of the hyperplane. [33] used SVMR to predict the daily maximum temperature and peak energy consumption of the heating system. [34] uses it to predict soil temperature. [35] improved SVMR by using Randomized Complete Block (RCB) and K-fold cross Validation techniques to predict temperature changes. [36] uses principal component analysis to assist SVMR to help aquaculture provide decision making. [37] combined the SVMR with the nearest neighbor neural network to predict water quality.

Least square support vector machines (LSSVM): LSSVM projects the samples into a high-dimensional feature space to make the samples linearly separable. Some scholars use LSSVN with a linear kernel function. [38]uses LSSVM with the linear kernel to predict dissolved oxygen content in crab culture. The radial basis kernel function is used for complex parameter relations, [39] proposed the LSSVM-GSA model for time prediction of wind power efficiency. [40] used environmental data from solar greenhouses to predict temperatures based on LSSVM. [41] used an improved group optimization LSSVM model to predict the dissolved oxygen content in crab culture. [42] used it to predict the concentration of dissolved oxygen.

K-mean algorithm support vector machine: The time series data is divided by a clustering method with different trends and then is input into the SVMR model. A comprehensive early-warning index system of water quality parameters based on SVMR was established by [43], which provides early warning for the aquaculture water environment. A prediction model of dissolved oxygen based on K-means was proposed by [44]. The mean absolute percentage error and root mean square error reach 1.4 % and 10.8 %, respectively. The K-mean algorithm support vector machine was improved by combining it with the Gated Recurrent Unit [45]. The mean absolute error of the new model is 0.264, and the mean absolute percentage error is 3.5.

Principal Component Analysis (PCA): PCA simplifies multivariable synthetically under the principle of minimizing data information loss. The dimensionality reduction algorithm makes a transparent relationship of parameters [45], uses principal component analysis to reduce parameter dimensions, and then uses GRU to build a prediction model of dissolved oxygen. [46] explored PCA combined with two post-processing techniques to predict solar irradiance. [47] improved cultural fish swarm algorithm based on PCA and LEAST-LSSVM was proposed to predict the pH value of aquaculture water quality.

3.2.3. *Model explores data relationships by itself*

In the advanced stage of data development, researchers use intelligent models to self-explore the relationship between greenhouse environmental parameters, usually using neural networks and their improvement methods.

Multilayer Perceptron (MLP): The Multilayer perceptron automatically searches for the relationship between parameters from the perspective of information processing. Due to its self-exploration, its accuracy is higher than that of traditional models. [48] used multi-layer perceptron to model the environment and compare with traditional models, three-layer MLP-type networks are already superior to other networks. [17] used multi-layer perceptron to train the perceptron model and predict greenhouse climate data.

Extreme learning machine (ELM): The Extreme learning machine improves multi-layer perceptron, whose connection weight w and the threshold value b are randomly generated and need not be adjusted. [49] used four kinds of extreme learning machines to predict water quality and dissolved oxygen concentration. The R-ELM has the best prediction accuracy. [50] compared the kernel function extreme learning machine with other models in the prediction of greenhouse environmental factors. The speed of the new method (0.0222 s) was faster than Back Propagation (0.7469 s) and SVM (19.2232 s) models and the accuracy was highest.

Bayesian Neural Network (BNN): The BNN integrates the idea of probability into the multi-layer perceptron, whose weight is that the random variables obey the Gaussian distribution. [51] uses BNN to connect measured data related to environmental control data to predict growth indicators. [52] used BNN to reduce aquaculture risk by reducing water quality predictions to between 0.5 % and 11 %.

Radial Basis Neural Network (RBNN): RBNN is based on the multi-layer perceptron, using a radial basis function as a hidden element to construct hidden layer space. [53] used RBNN to predict the environmental temperature. [54] used K-means and subtraction clustering to improve the radial basis neural network, accurately predicting the change of dissolved oxygen content in crab ponds.

3.3. *For feature Spatio-temporal correlations*

Multi-layer perceptrons can self-explore the complex relationships between environmental parameters, but they lack effective use of time information in the field of time series. This chapter systematically summarizes the deep neural network that can grasp the temporal and spatial correlation.

Recurrent neural network (RNN): RNN stores past information and current input by introducing latent variables, capturing the relationship between parameters and the space–time correlation of parameters. According to Fig. 3.

In the environment of facility agriculture, each parameter has a strong correlation and depends on the change of the last period. RNN can effectively capture the Spatio-temporal correlation of parameters. [55] introduced RNN to control the environment of agricultural facilities. In this model, artificial control variables are innovatively introduced into the model. With the upgrading of algorithms and the appearance of controllers with better performance, more optimized RNN environmental prediction application cases of agricultural facilities have appeared. [56] uses RNN combined with residual block, [57] used residual block to preprocess and extract representative features. The residual block sequence of the TCN model is used to process elements and finally input them into the fully connected network.

The MSE of the three data sets was controlled at 10.45 ± 0. 94, 6.76 ± 0.45, and 7.40 ± 1.88, respectively, which were the lowest among all methods. [58] used RNN to predict long-term power load at one-hour sampling frequency and analyzed models' relative performance with different power consumption patterns. RNN is used to predict the power consumption within the time range of one week for the building indoor power consumption at the sampling frequency of one hour, which improves the accuracy of the previous MLP model, as shown in Fig. 4.

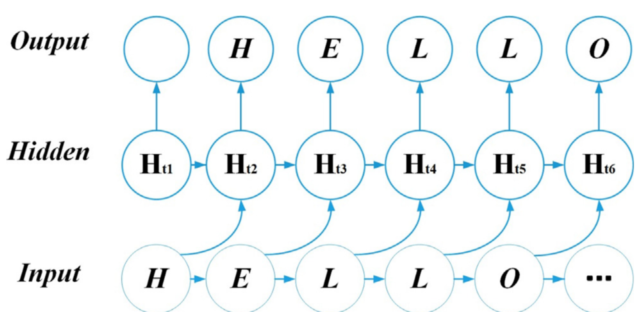

Fig. 3 – Model theory of RNN [55].

The relative error of the RNN and MLP models as a function of the root mean square average of total consumption per hour.

Long- and short-term memory neural network (LSTM): A single hidden layer of RNN cannot preserve long-term information and short-term input jump. The inventor of LSMN [59] introduced three gates and a candidate memory unit into the memory unit, forget gate, input gate, output gate, and candidate memory unit. The update mode is shown in Fig. 5.

The LSTM model can accurately predict periodic temperature or humidity fluctuation by forgetting or inheriting the state of the last time by iterating the data parameters, which makes fitting the nonlinear parameters easier. [60] monitor greenhouse climate using wireless sensor networks and LSTM to predict environmental parameters, using AME, MSE, and RMSE values as an evaluation index. LSTM combined with linear interpolation (LI) and random oversampling (OS) to predict local precipitation in the vicinity of the greenhouse in the next hour, providing reference results for greenhouse farmers and carrying out intelligent greenhouse control[61]. The recall rate increased from 47 % to 83 %. A more customizable four-season LSTM model was also made to match the Local Precipitation Forecast better. The rainfall pattern in different seasons has the highest accuracy of 96.20 %. Compared with other traditional models, LSTM has the highest precision and the lowest data fluctuation. The comparison between ANN, NARX, and RNN-LSTM showed that the overall prediction accuracy of RNN-LSTM was the highest. Detailed measurement results are compared in Fig. 6[62].

Gated recurrent unit (GRU): LSTM solves the more complex nonlinear problem in time series prediction, but its disadvan-

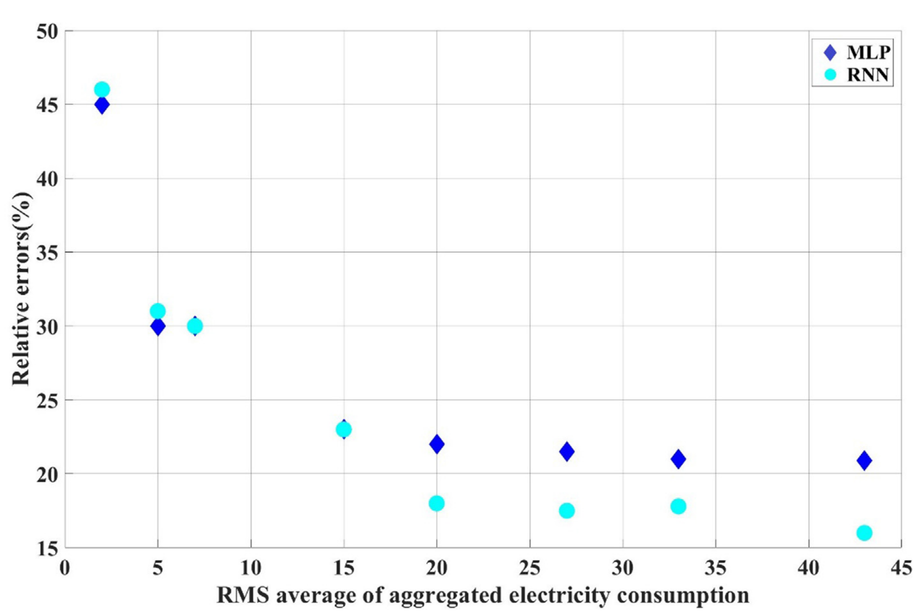

Fig. 4 – Accuracy of the previous MLP and RNN [58].

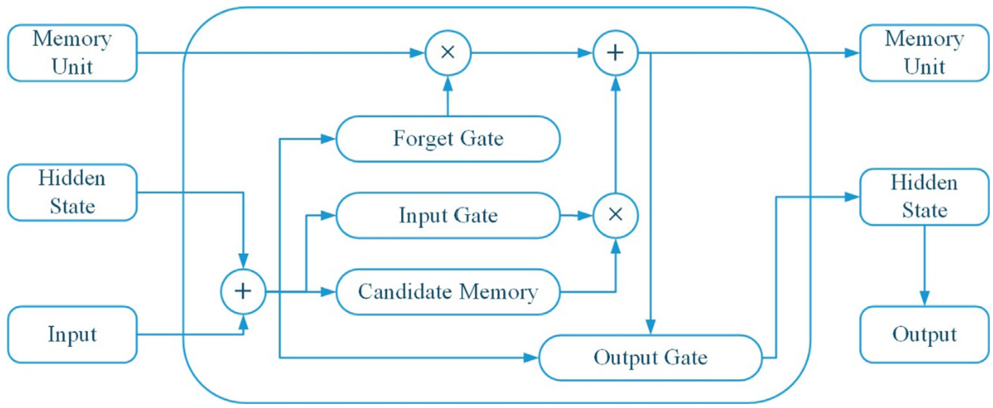

Fig. 5 – Model theory of LSTM [60].

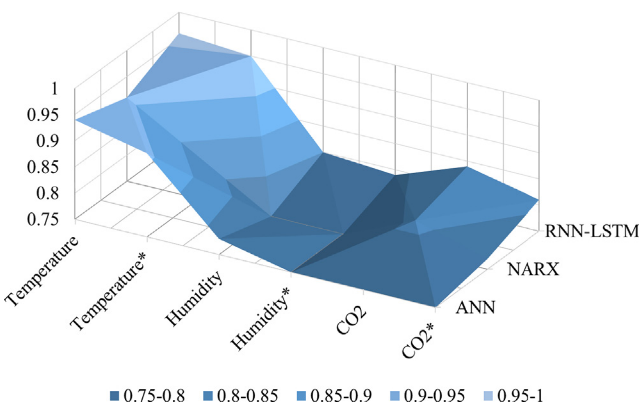

Fig. 6 – Comparison of the determination coefficient R^2 of ANN, NARX, and RNN-LSTM models for predicting different environmental parameters [62], where "*" indicates that experimental data excludes the influence of outdoor temperature, while "*" indicates that the influence of outdoor temperature is excluded.

tages are long training time, many parameters, and complex internal calculations. As a variant of LSTM, GRU was proposed in 2014 to simplify LSTM while keeping its effect unchanged. Many scholars replaced LSTM with efficient and straightforward GRU. [63] designed a multi-point temperature and humidity prediction method for mushroom houses based on the combination of CNN and GRU, and constructed the distribution characteristics by using historical meteorological data of greenhouse.

As Table 3 indicates, compared with LSTM, GRU has less running time, and the effect is similar. CNN-GRU further reduces the running time. [45] proposed a prediction model for dissolved oxygen in pond farming based on a combination of K-mean clustering and GRU neural network, which can predict the parameters of aquaculture water at different time intervals according to the requirements of actual scenarios. In the sample interval of 30 min, the mean absolute error is reduced to 0.264, and the mean absolute percentage error is 3.5 %.

Encoder-Decoder structure: The Autoencoder model was proposed by [64]. It is composed of an encoder and a decoder for high-dimensional complex data processing. As Fig. 7, the

Table 3 – Loss and time comparison of models [63].

Model	Training time(s)	Loss
BP	140.64	0.0022
LSTM	377.77	0.0020
GRU	360.38	0.0020
CNN–GRU	263.37	0.0028

seq2seq model is widely used in the field of translation. Its encoder and decoder are RNN, used to read input sentences and output translations, respectively.

This framework is also effective in the complex environmental prediction of agricultural facilities. The autoencoder [65] can compensate for errors and missing values in environmental parameter monitoring and reconstruct the time series of indoor environmental data. The three kinds of neural networks were constructed to detect multiple locations and parameters, respectively. Better than the polynomial interpolation model, this model effectively fills the gap. The environmental data to the data of indoor environment hour forecast also has higher precision.

3.4. *For complete Spatio-temporal information*

The deep learning model can grasp the Spatio-temporal correlation of parameters and improve data utilization. However, the grasp of long-term forecasts and irregular structure information needs to be improved. The attention mechanism and Graph neural network solve these problems and represent the development direction of deep learning.

Attention mechanism: The basic principle of the Attention mechanism can be described as a mapping from a query to m key-value pairs, simplified as Fig. 8.

In the application of agricultural facilities, two attention-based methods(Temporal-Attn, Spatiotemporal-Attn) have been proposed and compared with many models of water environment. Their comprehensive performance is better than other traditional methods. The results show that the Attention mechanism can reduce the influence of time steps on accuracy [66]. [67] proposed an attention-based 2D multi-input LSTM model, which combined DA-RNN [68], 2D-LSTM

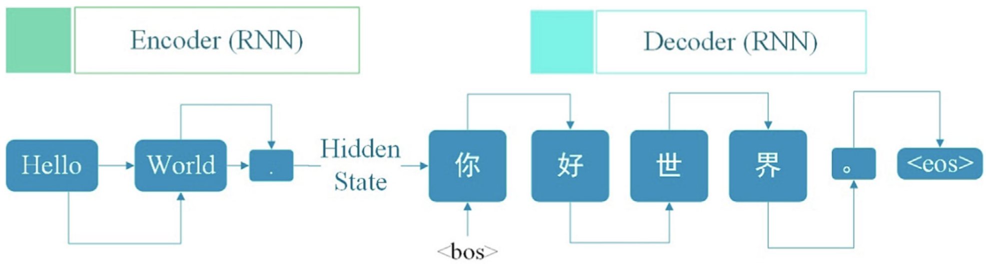

Fig. 7 – An example of the Encoding-Decoder model [65].

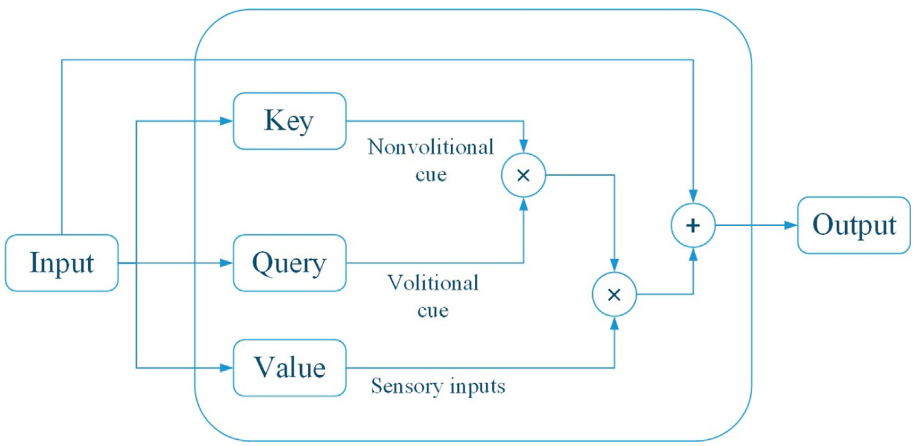

Fig. 8 – The basic principle of the Attention mechanism.

[69], and MI-LSTM [70]. H-attention and T-attention can capture the potential information of output values, while T-attention can capture more time information. The structure of the modified model can be briefly expressed as Fig. 9.

Graph Neural Network (GNN): GNN improves the deep learning model by complementing the structure information and node information in the environment and optimizing the experimental prediction task end-to-end. It can be used to realize information fusion that can explain artificial intelligence [71], which has excellent potential, and it helps manage cities by monitoring traffic flows [72]. It was developed as a diagnostic tool for chest CT images [73] and a spread prediction tool [74] during the COVID-19 pandemic, showing its ability to capture information. In environmental prediction, geographic information is usually predefined before training and input into the GNN model as structural information. A dynamic GNN was proposed by [75], which can predict soil moisture by using the dependence of relevant positions. [76] uses the PGNN models to effectively apply physical induced deviations to simulate the interactions between wind turbines. The GNN model proposed by [77] can accurately predict farm wind speed with clustering characteristics. The GNN-based model [78] proposes constructing the complex spatial dependence relationship of surface water quality sites to realize the multi-step prediction of surface water quality indicators. This indicates that the model can be further applied to the environmental parameters of agricultural facilities, explore the relationship of the parameters in different regions, and provide multi-temporal regional linkage decision-making suggestions for management.

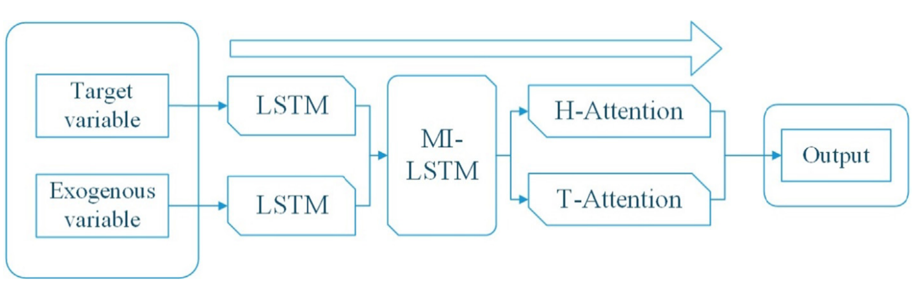

Fig. 9 – Examples of Attention mechanisms [67].

4. Discussion

4.1. *General discussion*

According to the evaluation scheme [24], searched a wide range of papers, narrowed the scope of the final papers, and conducted comprehensive quantitative research and analysis. At present, the validity and rationality of the study need to be further confirmed, and externalities or structural problems were explored according to the research methods provided by [79]. In article search, we combine complex strings to find articles, perform auditable screening, and obtain the repeatability of the final article, which has a fair and reasonable prospect for the research topic.

In the process of system review, the evaluation indicators of the paper are collected and quantified to evaluate the accuracy and generalization of the review conclusions. Some articles did not specify the evaluation indicators of the experimental results but compared the predicted results with the actual results in the form of charts. This may affect the reliability of the review conclusions. Some papers do not clearly state the source of the experimental data and the location of the model application, whether it is in the laboratory or in the factory, which will affect the generalization evaluation of the model.

Prior to analyzing each model in-depth, we analyze previous studies in the industry and list several flaws that will be covered in this review analysis: There is no discussion of the model's use of the internal organization of greenhouse environmental data, generalization of the model to the production environment, application of the time series model to actual production from the standpoint of the model's development process, or in-depth discussion of parameter setting.

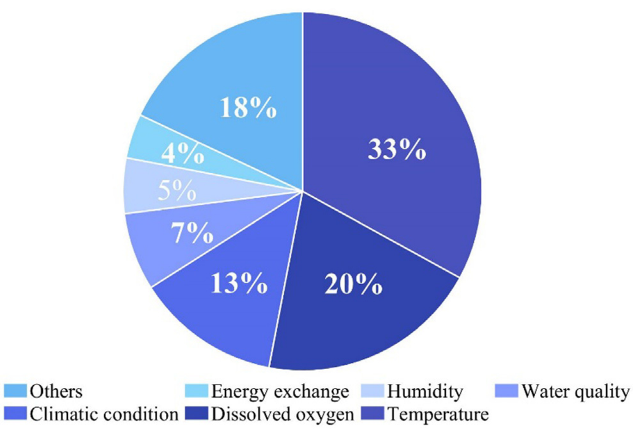

Fig. 10 – Detection target statistics.

4.2. *Problem discussion*

RQ1-Related (parameters) discussion: We focused on the impact of agro-environmental elements on model performance after a qualitative evaluation of model complexity and application impacts. These approaches primarily forecast the temperature (32.6 %) and dissolved oxygen (19.6 %) concentration of the environment, as shown in Fig. 10, since temperature and oxygen content strongly influence organisms. Temperature is essential in all types of greenhouses. The cultivation tank is an integral part of the greenhouse, and detecting dissolved oxygen is the key to monitoring. In addition, more models will choose the temperature, dissolved oxygen, and humidity as parameter properties in multivariable time series. According to the data correlation analysis, these characteristics influence the prediction goal, as shown in Fig. 11. When considering the biological effects of parameters, the temperature is the most widely studied, and other factors such as greenhouse ventilation, humidity, and the content of dissolved oxygen in aquaculture tanks are usually also related to the study of biological conditions.

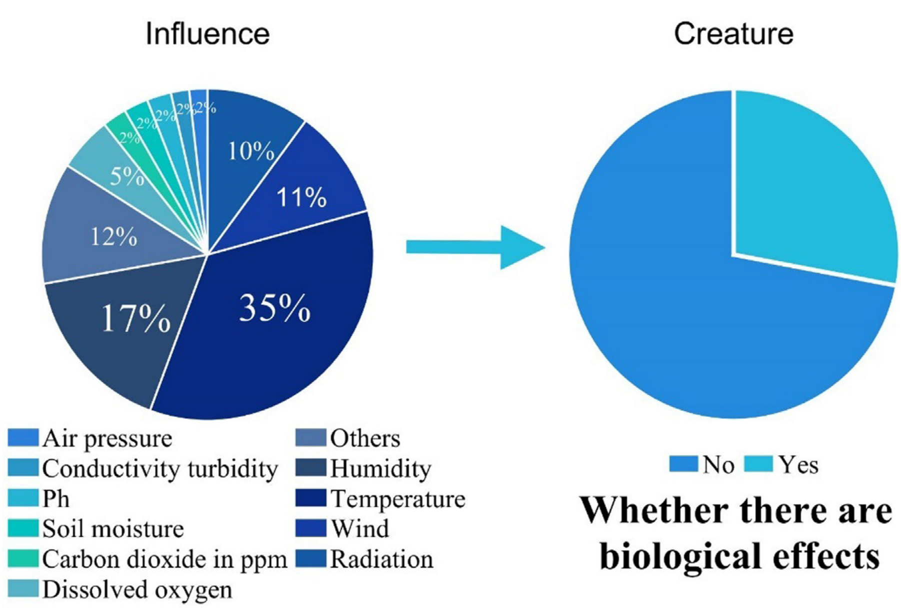

Fig. 11 – Environment parameter selection statistics.

With the continuous upgrading of the model, high requirements are put forward for the richness of parameters, which objectively promotes the mining of unknown variables and the increase of data volume. Set new environmental parameters, human factors, or biological factors as the detection target or input, such as wind speed, radiation, and wind direction, to increase the comprehensiveness of the input information. More comprehensive data can also improve or facilitate the development of models. For example, geographic information promotes the application of Graph neural networks in agricultural facilities.

RQ2-Related (accuracy and reliability assessment) discussion: According to statistics in Fig. 12, the parameters to test the model's accuracy are mostly MAE, MSE, R^2, or their deformation RMSE. There are also particular detection indicators such as accuracy, TIC, SD, and unique experiments.

- Small step size for high accuracy: Several studies have shown that the accuracy of machine learning models' predictions of environmental parameters is inversely associated with the interval time between measurements. When the time step is raised from 5 to 30 steps in the two datasets of the DSTP-RNN experiment [80], the RMSE of the least adaptable model GeoMAN grows by 190 % and 30 %, respectively. The best performing DSTP-RNN model has an RMSE rise of less than 5 % within 30 steps. However, when the step size exceeds 30 steps, the accuracy of all models increases exponentially in a catastrophic manner.

- Effective model structure improves model's stability: The RNN and LSTM memory modules improve the time series model's memory for periodic environmental changes, while the attention mechanism improves the model's capacity to simulate extended sequences by simulating human attention. To optimize the network structure, the DA-RNN model includes an attention mechanism [68]. The accuracy increases by 16 % when moving from three to ten steps and only by 5 % when moving from 10 to 25 steps. Compared to the old model, the error grows dramatically when the step size increases exponentially. This suggests that attention-based models can be prioritized in the future for predicting agro-environmental factors, which merits additional investigation and testing in facility agricultural applications.

RQ3-Related (data utilization) discussion: Based on a new perspective on information utilization, the application of advanced algorithms in agricultural facilities is reclassified and discussed. The internal structure of the model is the starting point for developing a new generation of time series models. Experiments have demonstrated that adding new feature structures to the model can help it acquire differentiating information, with varying degrees of improvement. A summary of the adequate information of existing models on the data can be summarized in Fig. 13.

- Neuron structure grasps the multifactorial information of the environment: For environmental prediction with little change in a single scene, autoregression has a high degree of precision. Spatial information is incorporated into the facility's agricultural setting by installing multi-layer perceptrons. [44] uses a neural network to reduce the SVM model MAPE from 7.41 % to 1.47 %. [51] uses a neural network to reduce the error rate of the regression model from 49.56 % to 30.13 %. The article believes that the neural network can spontaneously consider the existing agronomic knowledge and improve the accuracy of prediction. Graph neural network introduces graph structure based on MLP, which can grasp more macro geographic information and increase model accuracy. [77] introduces graph structure to reduce MAE by 14.1 % and 16.5 % in different step length experiments respectively.
- The memory structure enables the model to utilize information in the time dimension: A memory module is created to capture the spatiotemporal correlation of facility agricultural ambient information, and the deep neural

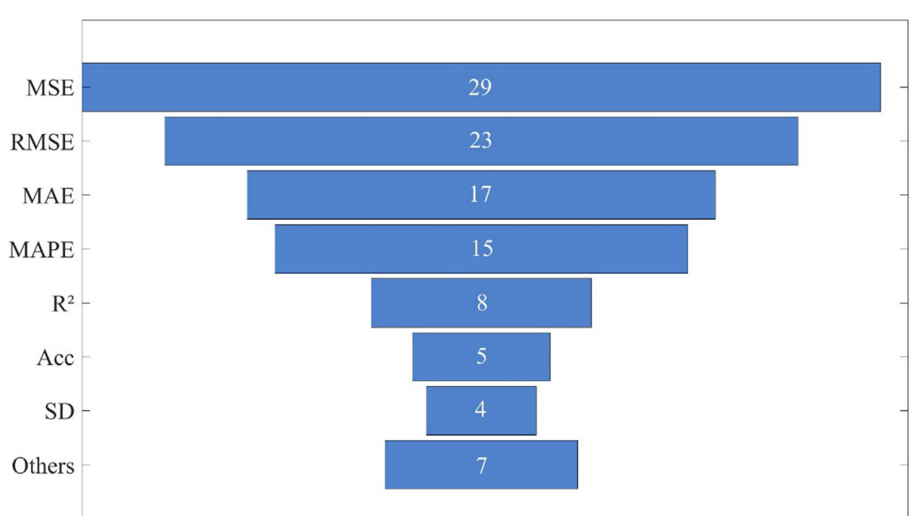

Fig. 12 – Evaluation index statistics.

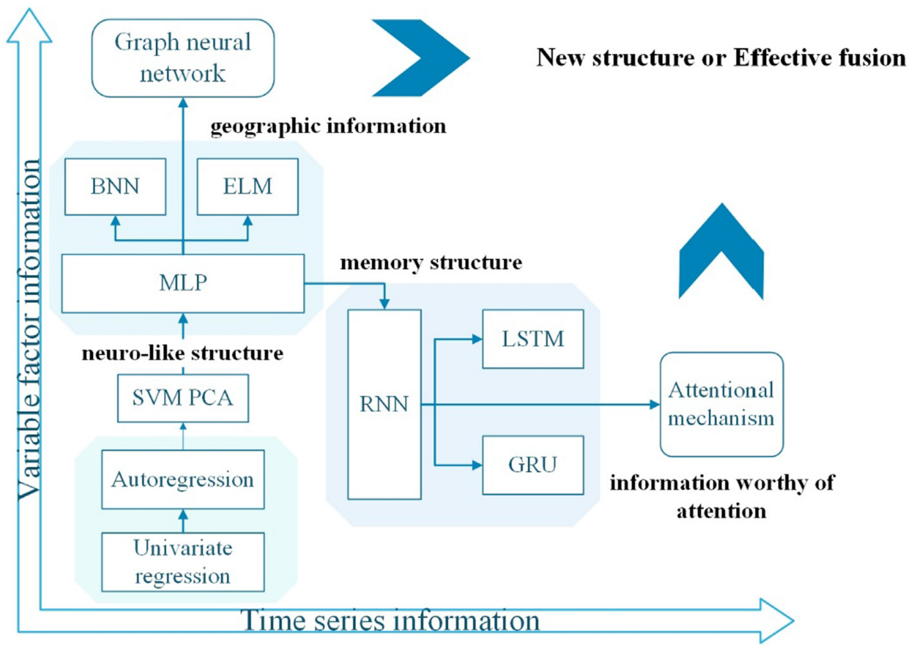

Fig. 13 – Utilization of two-dimensional data in time series models.

network (RNN, LSTM, etc.) is based on the superposition of multi-layer perceptrons. In a facility agricultural environment, investigating temperature and dissolved oxygen periodicity. [45] uses GRU to reduce the MAE of CNN without temporal module by 35.3 % and ELM by 27.9 %. In the experiment [80], the MAE of the LSTM model is reduced by 29.5 % and 5.0 % compared with the autoregressive model. After the attention module is introduced, the Attn-RNN is reduced by 49.6 % and 60.8 % compared with the LSTM. The initial tests suggest that adding a memory module to a neural network makes it easier to learn the spatiotemporal correlation of the greenhouse environment and improves the capacity to interpret extended sequences.

- The model's application and data collection complement each other: With the improvement of computer power and the passage of time, algorithms for predicting environmental parameters of greenhouse facilities are also constantly evolving. The upgrade of the new model in different periods is as follows: The examination of the model's data structure continues to be more profound as the data gets more diversified. Researchers used autoregression to investigate spatiotemporal connections between parameters and then introduced spatiotemporal modules. To accomplish end-to-end optimization, the latest models can comprehend overall data changes or examine data structure information from many dimensions. Simultaneously, the upgraded algorithms impose increased data volume and variety requirements, forcing researchers to gather more and various data types.

RQ4-Related (forward-looking) discussion: The agricultural field is characterized by low timeliness, long experiment time, and difficulty in data collection. In the environment of agricultural facilities, both the temperature change of one day or one year in thermal environment prediction and the change of dissolved oxygen content in the water environment have periodic and complex nonlinear characteristics. Different parameters in the environment also have strong temporal and spatial correlation, which presents a tough challenge for the simulation of a complex environment. However, as illustrated in Fig. 14, researchers have been actively applying Deep Learning models to predict the environmental parameters of agricultural facilities. We make a brief conclusions about the research trends.

- **Time series modeling hotspot turns to deep learning models:** As shown in Fig. 14, the autoregressive and traditional machine learning models have an extended period, and the years are concentrated in 2015 and 2016, while the average value of methods such as deep learning is concen-

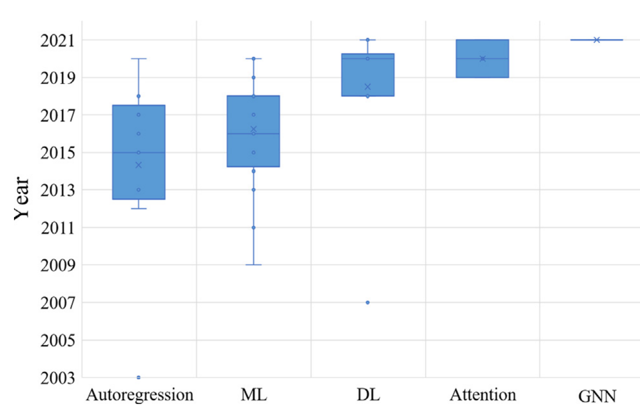

Fig. 14 – Year distribution of time series model application in the greenhouse environment.

trated around 2020. To show trends, we separate attention-enhanced deep learning models or Graph neural networks. It can be seen that the attention mechanism has almost become the standard configuration of the SOTA model after 2019.

- **The specific structural features of the deep model need to be further explored:** RNN, LSTM, and other models use specific structures to extract temporal and spatial correlations. The Attention mechanism grasps the overall situation and improves the interpretability of the relationship between environmental parameters. The Graph neural network mining deep data structure from a multi-dimensional perspective. To a certain extent, the Graph neural network is rarely used in environmental prediction and only gives structural information to nodes through predefined geographic information. In the future, multiple structures between parameters can be further explored. Fig. 15 summarizes the relationship between time step selection and accuracy, and it could be demonstrated that MAEs of spatio-temporal-attn, temporal-attn, and LSTM are minimal compared with others corresponding to smaller intervals. However, for bigger intervals, MAEs of spatio-temporal-attn and temporal-attn are minimal. Observed from Fig. 15, it can be inferred that the accuracy of the model lacking the attention mechanism, such as MLP and LSTM, would drop off quickly when the time interval setting of the experiment grows. This indicates that the attention mechanism can significantly mitigate the impact of the detection time interval on the experimental accuracy.

4.3. Horizontal model comparison

After a systematic literature review and generalization of the model through the model's effectiveness, we combined the actual application of the model in the facility agricultural environment and compared the model horizontally from the level of model principles.

The traditional autoregressive model (ARMAX) and the model of the primary stage of machine learning (Multivariate regression) are similar in application, using a target parameter as the primary variable. The main function is used to simulate the trend of the target variable. Other influencing factors are appropriately added for precision fitting based on the target parameters. Their advantages are simple models, few parameter choices, and quick application deployment. The disadvantage is that the data utilization efficiency is low, and the structure can only be affected at the numerical level. The intermediate stage of machine learning application (SVMR, PCA) maps the data to a high-dimensional space, makes the coupled data information linearly separable, replaces the complex high-dimensional mapping with the inner product kernel function, and then finds the largest split surface through the dimensionality reduction method. Finally, realize the prediction of multi-source data. Although facility agriculture forecasting has increased the forecast accuracy several times, the target parameter monitoring cycle is long in a complex agricultural environment, and the relationship between environmental parameters is complicated, making SVMR unable to handle it. When evaluating utilization efficiency, its model uses the spatial relationship between the parameters at the physical level but does not explore the internal connection between the data and loses much information, resulting in poor generalization in the facility's agricultural environment.

Starting from the advanced stage of machine learning, the multi-layer perceptron introduces neurons and uses artificial neural networks to fit the complex relationships between parameters. The Deep neural networks (RNN, LSTM, etc.) are based on multi-layer perceptron stacking and introduce memory modules to capture the spatiotemporal correlation

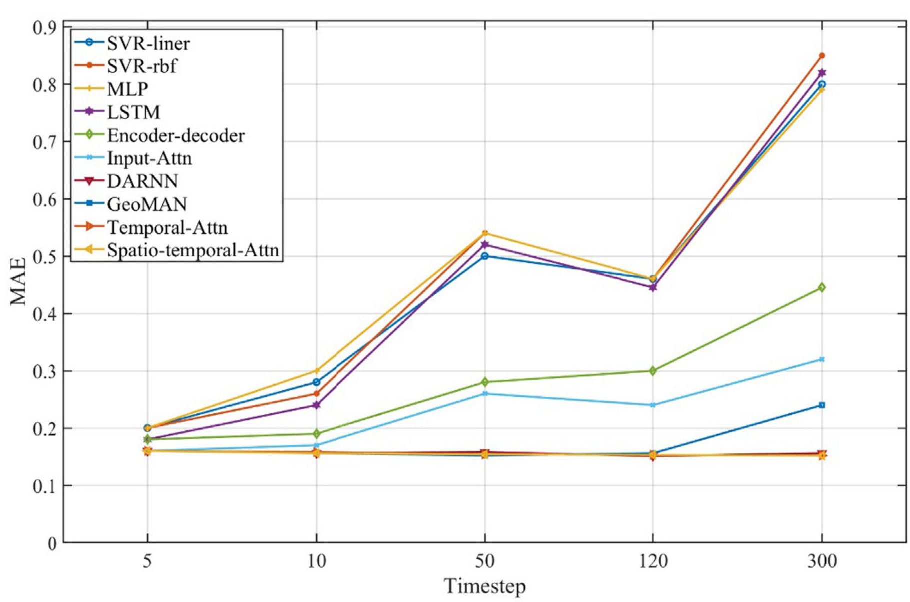

Fig. 15 – Prediction accuracy of different models at different time interval choices [66].

of information in the facility's agricultural environment. In the environment of facility agriculture, temperature and dissolved oxygen often have periodic laws, while other factors are often inherently related to the target parameters. The practical structure of the deep neural network can accommodate the massive data in the greenhouse environment, simulate the data structure, improve the information utilization rate, and raise the greenhouse modeling to a higher level. The latest research direction of deep neural networks in facility agricultural environment modeling is the Attention mechanism and Graph neural network. The facility agriculture environment requires long-term prediction and the overall grasping ability. The Attention mechanism can grasp the complete Spatio-temporal correlation, accurately grasp the complete relationship of the parameters, and achieve high-precision prediction of the facility's agricultural environment. At the same time, a practical model similar to the Attention mechanism shows cross-domain generalization in natural language processing, and it needs to be further applied to facility agriculture. The researcher's structural characteristics of sensor placement and environmental parameters in facility agriculture form a multi-agent and multi-attribute topological map. The graph of the Graph neural network represents learning characteristics, which can be well integrated with the environmental topological map to construct a Complete Spatio-temporal correlation. At present, researchers only introduce geographic information into the Graph neural network in predicting the environment of facility agriculture and do not define the relationship between the attributes. Although it is possible to grasp the macro characteristics of the data at the three-dimensional level, the utilization efficiency of the parameters still needs to be improved.

4.4. *Limitations*

This review mainly reviews the application of time series models in the greenhouse field and comprehensively evaluates the advantages and disadvantages of the models. Although it is objectively evaluated according to the SLR principle, it is still highly subjective. Second, this review focuses on time series models based on deep learning, neglecting traditional machine learning and statistical time series models.

5. Conclusions

This article reviews the development history of applying artificial intelligence-based time series models in horticultural environments. From the perspective of environmental information utilization, a discussion based on the accuracy and generalization of the model was launched, and the influence of hyperparameters such as environmental parameter selection, detection environment time interval, model evaluation index, and other hyperparameters on the model application was discussed. Deep learning has expedited artificial intelligence modeling, which is reflected in facility agricultural time series modeling, and the SOTA model is regularly updated. However, present models are limited in their adaptability in agricultural facility applications, as only 21 % of articles have adequate generalization rates and can only be based on a specific experimental setting.

As predictive models progress from statistical or machine learning models to deep learning models, we can conclude that models' ability to mine data grows, and the ability to use additional data drives models in more complicated directions. Temperature and dissolved oxygen parameters in the surroundings of agricultural facilities are the most considered parameters, and the model's complexity makes parameter prediction more accurate. The summary of the models reveals that the deep learning model performs better than the models based on statistical and machine learning. The model's setting shows that prediction accuracy declines as the forecast time step increases. Nonetheless, attention-based models can greatly offset the effect of improving prediction accuracy with time increments. However, current research indicates that graph neural networks can be effectively combined with environmental topological maps to create complete Spatio-temporal connections. However, graph neural networks are seldom employed for environmental prediction of the agricultural greenhouse.

In the future, intelligent deep neural network models will be explored further, focusing on the global grasp ability of attention mechanisms and the structural information grasp ability of Graph neural networks and to deepen the understanding of the accuracy and generalization of prediction tasks.

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 research was financially supported by Overseas High-level Youth Talents Program (China Agricultural University, China, Grant No. 62339001), Science and Technology Cooperation – Sino-Malta Fund 2019: Research and Demonstration of Real-time Accurate Monitoring System for Early-stage Fish in Recirculating Aquaculture System (AquaDetector, Grant No. 2019YFE0103700), China Agricultural University Excellent Talents Plan (Grant No. 31051015), Major Science and Technology Innovation Fund 2019 of Shandong Province (Grant No. 2019JZZY010703), National Innovation Center for Digital Fishery, and Beijing Engineering and Technology Research Center for Internet of Things in Agriculture. The authors also appreciate constructive and valuable comments provided by reviewers.

Appendix

Summary of Machine Learning algorithms.

Group[a]	Publication/reference	Method[b]	Year	Influence factors and sample design for modeling[c]			
				Targets	Interval	Outside climates	Inside climates
Explore the self-variation of indicators	Hui et al.	ARMAX	2017	Temperature	1 min	outside temperature, outside humidity, solar radiation, wind speed	inside temperature
	Ma et al.	NNARMAX	2015	Humidity	1 min	outside temperature, humidity, wind speed, wind direction, solar radiation	inside temperature
	Guzmán-Cruz et al.	ARMAX	2013	Temperature	5 min	outside temperature, humidity, wind speed, solar radiation	N/A
	Frausto et al.	ARMAX	2003	Temperature	5 min	outside temperature, humidity, wind speed, solar radiation, cloudiness of the sky	N/A
	Xu et al.	IARX	2016	Temperature	15 min	outside temperature, humidity, wind speed, solar radiation, cloudiness of the sky	N/A
	Gustin et al.	ARX	2018	Temperature	15 min	outside temperature, solar irradiation	inside temperature
	Xiong et al.	IARX	2012	Temperature	1 min	outside temperature, humidity, solar radiation	inside temperature
	Tian et al.	MA-ARIMA-GASVR	2020	Temperature	1 h	outside temperature, humidity, wind speed, solar radiation	greenhouse structural, cover materials, ventilation
For exogenous factors influence	García García et al.	Multivariate Regression	2019	LCA	90 days	abiotic depletion, global warming, ozone layer depletion	N/A
	Fang, Tingting, et al.	Multivariate Regression	2016	Heat demand	72 h	outside temperature	inside temperature
	Paniagua-Tineo et al.	SVMR	2011	Temperature	N/A	global radiation, synoptic situation, sea level pressure,	relative humidity, precipitation
	Xing et al.	SVMR	2018	Temperature	N/A	relative humidity, wind speed, rainfall, solar radiation	air temperature, soil temperatures,
	Taki, Morteza, et al.	SVMR	2016	Energy exchange	1 h	outside temperature, wind speed, solar radiation	inside soil temperature, inside air humidity
	Shahriar et al.	SVMR	2014	River flow	1 day	rainfall	N/A

Group[a]	Publication/reference	Method[b]	Year	Influence factors and sample design for modeling[c]			
				Targets	Interval	Outside climates	Inside climates
Aldhyani et al.	SVMR	2020	Temperature	N/A	global radiation, synoptic situation, sea level pressure	relative humidity, precipitation	
Liu et al.	LSSVM	2014	Dissolved oxygen	20 min	N/A	N/A	
Yuan et al.	LSSVM	2015	Wind power	1 h	N/A	N/A	
Yu, Huihui, et al.	LSSVM	2016	Temperature	1 h	outside temperature, wind speed, solar radiation	air humidity, soil temperature, soil moisture	
Liu et al.	LSSVM	2013	Dissolved oxygen	30 min	solar radiation, wind velocity	water temperature, rainfall, humidity, dissolved oxygen	
Li et al.	LSSVM	2018	Dissolved oxygen	20 min	temperature	PH, salinity, conductivity turbidity	
Liu et al.	K-mean SVM	2018	Water quality	1 h	atmospheric humidity, atmospheric temperature	water temperature, dissolved oxygen, acidity	
Huan et al.	K-mean SVM	2016	Dissolved oxygen	15 min	air pressure, air temperature, wind speed	water temperature, PH	
Davòet al.	PCA	2016	Dissolved oxygen	30 min	air pressure, wind speed, solar radiation	conductivity, temperature, dissolved oxygen, PH	
Liu et al.	PCA	2014	PH	10 min	outside temperature	PH, dissolved oxygen, water temperature, oxygen, reduction potential	
Francik et al.	Multilayer Perceptron	2020	Temperature	1 h	outside temperature	moisture, internal temperature	
Dariouchy et al.	Multilayer Perceptron	2009	Climatic conditions	5 s	temperature, humidity	soil temperature, internal moisture	
Heddam et al.	ELM	2017	Dissolved oxygen	1 h	temperature	PH, dissolved oxygen, turbidity	
Liu et al.	ELM	2016	Climatic conditions	1 h	temperature, humidity	dissolved oxygen	
Kocian et al.	BNN	2020	Leaf Area Index	1 h	irradiance	temperature	
Dabrowski et al.	BNN	2018	Water quality	15 min	N/A	dissolved oxygen, PH	
Taki et al.	RBNN	2018	Energy exchange	1 h	outside temperature, wind speed, solar radiation	temperature, mean leaf width	
Chen et al.	RBNN	2018	Dissolved oxygen	1 h	rainfall, wind speed, solar radiation, air temperature	solar radiation, air temperature, water temperature	
For feature Spatio-temporal correlations	Fourati et al.	RNN	2007	Temperature	2 min	external temperature, external hygrometry, wind speed, global radiant	internal temperature, internal hygrometry

Group[a]	**Publication**/reference	**Method**[b]	**Year**	**Influence factors and sample design for modeling**[c]			
				Targets	**Interval**	**Outside climates**	**Inside climates**
Hongkang et al.	RNN	2018	Climatic conditions	10 min	outside temperature, illumination	air humidity, air temperature, concentration recorded	
Gong et al.	RNN	2021	Crop Yield	1 day	1 day	relative humidity, radiation	
Rahman et al..	RNN	2018	Electricity consumption	1 h	dry-bulb temperature	relative humidity, concentration recorded, temperature, humidity	
Ali, Asmaa, et al.	LSTM	2020	Climatic conditions	1 day	dew point, wind power	temperature humidity, air pressure	
Hsieh et al.	LSTM	2020	Instant heavy rainfall	1 h	humidity radiation	temperature	
Jung et al.	LSTM	2020	Climatic conditions	5 min	temperature	temperature, humidity, carbon dioxide	
Cao et al.	GRU, K-mean SVM	2020	Dissolved oxygen	30 min	air pressure, wind speed, solar radiation	conductivity, temperature, dissolved oxygen, PH,	
Zhao Quanming	GRU	2020	Humidity	5 min	illumination intensity	air temperature, relative humidity, soil temperature	
Liguori et al.	RNN Autoencoder	2021	Climatic conditions	30 min	temperature	temperature, humidity, carbon dioxide	
For complete	*spatiotemporal information* soil moisture, temperature	Liu, Yeqi, Zhang	Spatio-temporal Attn, Temporal-Attn	2019	Dissolved oxygen	10 min	meteorology, soil moisture, temperature
	Liu, Yeqi, et al.	DSTP-RNN	2020	Temperature	1 min	temperature, carbon dioxide	temperature, carbon dioxide in ppm, relative humidity
	Kim et al.	2DA-MILSTM	2021	Temperature	1 min	weather forecast temperature, carbon dioxide	temperature, carbon dioxide, relative humidity
	Xu Jiahui et al.	GNN	2021	Water quality	4 h	water quality	N/A
	Anoushka Vyas et al.	GNN	2020	Soil Moisture	15 days	soil moisture	N/A
	Chen et al.	GNN	2021	Wind speed	5 min	wind speed, wind direction	N/A
Park et al.	GNN	2019	Power estimation	N/A	wind speed	N/A	

Note:
a:“ Group” shows the classification of models according to the utilization of data and the depth of mining.
b:“ Method” describes the models or algorithms used to predict environmental parameters of agricultural facilities.
C:“ Influence factors and sample design for modeling” lists the research objectives and influencing factors in environmental parameter prediction of agricultural facilities, including time step selection and indoor or outdoor parameter selection.

REFERENCES

[1] Guo Y, Zhao H, Zhang S, Wang Y, Chow D. Modeling and optimization of environment in agricultural greenhouses for improving cleaner and sustainable crop production. J Clean Prod 2021;285. https://doi.org/10.1016/j.jclepro.2020.124843 124843.
[2] United Nations, Department of Economic and Social Affairs, Population Division, World population prospects Highlights, 2019 revision Highlights, 2019 revision, 2019. http://creativecommons.org/licenses/by/3.0/igo/.
[3] Mirzamohammadi S, Jabarzadeh A, Shahrabi MS. Long-term planning of supplying energy for greenhouses using renewable resources under uncertainty. J Clean Prod 2020;264. https://doi.org/10.1016/j.jclepro.2020.121611 121611.
[4] Saiz-Rubio V, Rovira-Más F. From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management. Agronomy 2020;10:207. https://doi.org/10.3390/agronomy10020207.
[5] Ramin Shamshiri R, Kalantari F, Ting KC, Thorp KR. Advances in greenhouse automation and controlled environment agriculture: A transition to plant factories and urban agriculture. Int J Agric Biol Eng 2018;11:1–22. Available from: https://doi.org/10.25165/j.ijabe.20181101.3210.
[6] A. Abdullah, S. Al Enazi, I. Damaj, AgriSys: A smart and ubiquitous controlled-environment agriculture system, in: 2016 3rd MEC Int. Conf. Big Data Smart City ICBDSC, IEEE, Muscat, 2016: pp. 1–6. https://doi.org/10.1109/ICBDSC.2016.7460386.
[7] S. Kamal, The Renewable Revolution: How We Can Fight Climate Change, Prevent Energy Wars, Revitalize the Ec... - Harvard Book Store, (2013). https://www.harvard.com/book/the_renewable_revolution_how_we_can_fight_climate_change_prevent_energy_war/ (accessed October 9, 2021).
[8] Engler N, Krarti M. Review of energy efficiency in controlled environment agriculture. Renew Sustain Energy Rev 2021;141. https://doi.org/10.1016/j.rser.2021.110786 110786.
[9] Elanchezhian A, Basak JK, Park J, Khan F, Okyere FG, Lee Y, et al. Evaluating different models used for predicting the indoor microclimatic parameters of a greenhouse. Appl Ecol Environ Res 2020;18:2141–61. Available from: https://doi.org/10.15666/aeer/1802_21412161.
[10] Z. Hui, Q. Lin-lin, W. Gang, Modeling and simulation of greenhouse temperature hybrid system based on ARMAX model, in: 2017 36th Chin. Control Conf. CCC, IEEE, Dalian, China, 2017: pp. 2237–2241. https://doi.org/10.23919/ChiCC.2017.8027690.
[11] R.R. Rhinehart, Nonlinear Regression Modeling for Engineering Applications: Modeling, Model Validation, and Enabling Design of Experiments, John Wiley & Sons, Ltd, Chichester, UK, 2016. https://doi.org/10.1002/9781118597972.
[12] Singh MC, Singh JP, Singh KG. Development of a microclimate model for prediction of temperatures inside a naturally ventilated greenhouse under cucumber crop in soilless media. Comput Electron Agric 2018;154:227–38. https://doi.org/10.1016/j.compag.2018.08.044.
[13] Choab N, Allouhi A, El Maakoul A, Kousksou T, Saadeddine S, Jamil A. Review on greenhouse microclimate and application: Design parameters, thermal modeling and simulation, climate controlling technologies. Sol Energy 2019;191:109–37. https://doi.org/10.1016/j.solener.2019.08.042.
[14] ZhiYu Z, HanPing M, XiaoDong Z, Jing H, LüHua H, Jing N. Forecast model of greenhouse temperature based on time series method. Nongye Jixie Xuebao Trans Chin Soc Agric Mach 2010;41:173–82.
[15] Song G, Dai Q. A novel double deep ELMs ensemble system for time series forecasting. Knowl-Based Syst 2017;134:31–49. https://doi.org/10.1016/j.knosys.2017.07.014.
[16] Chou C-K, Lo C-C, Huang S-H, Horng M-F, Kuo Y-H, Based GCC, et al. IEEE World Congr. Serv. Serv. IEEE, San Francisco, CA 2018;2018:61–2. https://doi.org/10.1109/SERVICES.2018.00044.
[17] Dariouchy A, Aassif E, Lekouch K, Bouirden L, Maze G. Prediction of the intern parameters tomato greenhouse in a semi-arid area using a time-series model of artificial neural networks. Measurements 2009;42:456–63. https://doi.org/10.1016/j.measurement.2008.08.013.
[18] B. Kitchenham, Guidelines for performing Systematic Literature Reviews in Software Engineering, (2007) 44.
[19] I.L. López-Cruz, E. Fitz-Rodríguez, R. Salazar-Moreno, A. Rojano-Aguilar, M. Kacira, Graduate Agricultural Engineering Program, University of Chapingo, Chapingo, Mexico, Agricultural and Biosystems Engineering Dept., University of Arizona, Tucson, USA, Development and analysis of dynamical mathematical models of greenhouse climate: A review, Eur. J. Hortic. Sci. 83 (2018) 269–279. https://doi.org/10.17660/eJHS.2018/83.5.1.
[20] Deb C, Zhang F, Yang J, Lee SE, Shah KW. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902–24. https://doi.org/10.1016/j.rser.2017.02.085.
[21] Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Comput Electron Agric 2018;151:61–9. https://doi.org/10.1016/j.compag.2018.05.012.
[22] van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. Comput Electron Agric 2020;177. https://doi.org/10.1016/j.compag.2020.105709 105709.
[23] Zhang S, Guo Y, Zhao H, Wang Y, Chow D, Fang Y. Methodologies of control strategies for improving energy efficiency in agricultural greenhouses. J Clean Prod 2020;274. https://doi.org/10.1016/j.jclepro.2020.122695 122695.
[24] B. Kitchenham, Guidelines for performing Systematic Literature Reviews in Software Engineering, (n.d.) 44.
[25] Frausto HU, Pieters JG, Deltour JM. Modelling Greenhouse Temperature by means of Auto Regressive Models. Biosyst Eng 2003;84:147–57. https://doi.org/10.1016/S1537-5110(02)00239-8.
[26] Xu Z, Chen J, Zhang J, Liu Y, Zhang Q. Incremental Auto Regressive Prediction Models with External Variables of Greenhouse Air Temperature for Control Purposes. Int J Smart Home 2016;10:45–58. Available from: https://doi.org/10.14257/ijsh.2016.10.9.05.
[27] Gustin M, McLeod RS, Lomas KJ. Forecasting indoor temperatures during heatwaves using time series models. Build Environ 2018;143:727–39. https://doi.org/10.1016/j.buildenv.2018.07.045.
[28] Guzmán-Cruz R, Olvera-González E, López-Cruz IL, Montoya-Zamora R. Modelling greenhouse air temperature using evolutionary algorithms in auto regressive models. Afr J Agric Res 2013:11. https://doi.org/10.5897/AJAR12.280.
[29] G. Ma, L. Qin, Z. Chu, G. Wu, Modeling greenhouse humidity by means of NNARMAX and principal component analysis, in: 27th Chin. Control Decis. Conf. 2015 CCDC, IEEE, Qingdao, China, 2015: pp. 4840–4845. https://doi.org/10.1109/CCDC.2015.7162782.
[30] D. Tian, Prediction of temperature in edible fungi greenhouse based on MA-ARIMA-GASVR, Trans. Chin. Soc. Agric. Eng. (2020). https://doi.org/10.11975/j.issn.1002-6819.2020.03.023.
[31] García García B, Rosique Jiménez C, Aguado-Giménez F, García García J. Life Cycle Assessment of Seabass

(Dicentrarchus labrax) Produced in Offshore Fish Farms: Variability and Multiple Regression Analysis. Sustainability 2019;11:3523. https://doi.org/10.3390/su11133523.

[32] Fang T, Lahdelma R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Appl Energy 2016;179:544–52. https://doi.org/10.1016/j.apenergy.2016.06.133.

[33] Paniagua-Tineo A, Salcedo-Sanz S, Casanova-Mateo C, Ortiz-García EG, Cony MA, Hernández-Martín E. Prediction of daily maximum temperature using a support vector regression algorithm. Renew Energy 2011;36:3054–60. https://doi.org/10.1016/j.renene.2011.03.030.

[34] Xing L, Li L, Gong J, Ren C, Liu J, Chen H. Daily soil temperatures predictions for various climates in United States using data-driven model. Energy 2018;160:430–40. https://doi.org/10.1016/j.energy.2018.07.004.

[35] Taki M, Ajabshirchi Y, Ranjbar SF, Rohani A, Matloobi M. Heat transfer and MLP neural network models to predict inside environment variables and energy lost in a semi-solar greenhouse. Energy Build 2016;110:314–29. https://doi.org/10.1016/j.enbuild.2015.11.010.

[36] Shahriar MS, Rahman A, McCulloch J. Predicting shellfish farm closures using time series classification for aquaculture decision support. Comput Electron Agric 2014;102:85–97. https://doi.org/10.1016/j.compag.2014.01.011.

[37] Aldhyani THH, Al-Yaari M, Alkahtani H, Maashi M. Water Quality Prediction Using Artificial Intelligence Algorithms. Appl Bionics Biomech 2020;2020:1–12. https://doi.org/10.1155/2020/6659314.

[38] Liu S, Xu L, Jiang Y, Li D, Chen Y, Li Z. A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. Eng Appl Artif Intell 2014;29:114–24. https://doi.org/10.1016/j.engappai.2013.09.019.

[39] Yuan X, Chen C, Yuan Y, Huang Y, Tan Q. Short-term wind power prediction based on LSSVM–GSA model. Energy Convers Manag 2015;101:393–401. https://doi.org/10.1016/j.enconman.2015.05.065.

[40] Yu H, Chen Y, Hassan SG, Li D. Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO. Comput Electron Agric 2016;122:94–102. https://doi.org/10.1016/j.compag.2016.01.019.

[41] Liu S, Xu L, Li D, Li Q, Jiang Y, Tai H, et al. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. Comput Electron Agric 2013;95:82–91. https://doi.org/10.1016/j.compag.2013.03.009.

[42] Li C. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. Inf Process Agric 2018:10.

[43] Liu Y. Research on Data Analysis Model of Aquaculture Based on Time Series Data. China Agricultural University; 2018.

[44] juan Huan, Dissolved oxygen prediction in water based on K-means clustering and ELM neural network for aquaculture, Trans. Chin. Soc. Agric. Eng. (2016). https://doi.org/10.11975/j.issn.1002-6819.2016.17.024.

[45] Cao X, Liu Y, Wang J, Liu C, Duan Q. Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network. Aquac Eng 2020;91. https://doi.org/10.1016/j.aquaeng.2020.102122 102122.

[46] Davò F, Alessandrini S, Sperati S, Delle Monache L, Airoldi D, Vespucci MT. Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. Sol Energy 2016;134:327–38. https://doi.org/10.1016/j.solener.2016.04.049.

[47] S. Liu, L. Xu, Forecasting Model for pH Value of Aquaculture Water Quality Based on PCA-MCAFA-LSSVM, Trans. Chin. Soc. Agric. Mach. (2014). 10.6041 /j.issn.1000-1298.2014.05.037.

[48] Francik S, Kurpaska S. The Use of Artificial Neural Networks for Forecasting of Air Temperature inside a Heated Foil Tunnel. Sensors 2020;20:652. https://doi.org/10.3390/s20030652.

[49] Heddam S, Kisi O. Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. Environ Sci Pollut Res 2017;24:16702–24. https://doi.org/10.1007/s11356-017-9283-z.

[50] Q. Liu, D. Jin, J. Shen, Z. Fu, N. Linge, A WSN-based prediction model of microclimate in a greenhouse using extreme learning approaches, in: 2016 18th Int. Conf. Adv. Commun. Technol. ICACT, IEEE, Pyeongchang Kwangwoon Do, South Korea, 2016: pp. 730–735. https://doi.org/10.1109/ICACT.2016.7423609.

[51] Kocian A, Carmassi G, Cela F, Incrocci L, Milazzo P, Chessa S. Bayesian Sigmoid-Type Time Series Forecasting with Missing Data for Greenhouse Crops. Sensors 2020;20:3246. https://doi.org/10.3390/s20113246.

[52] J.J. Dabrowski, A. Rahman, A. George, S. Arnold, J. McCulloch, State Space Models for Forecasting Water Quality Variables: An Application in Aquaculture Prawn Farming, in: Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM, London United Kingdom, 2018: pp. 177–185. https://doi.org/10.1145/3219819.3219841.

[53] Taki M, Abdanan Mehdizadeh S, Rohani A, Rahnama M, Rahmati-Joneidabad M. Applied machine learning in greenhouse simulation; new application and analysis. Inf Process Agric 2018;5:253–68. https://doi.org/10.1016/j.inpa.2018.01.003.

[54] Chen Y, Yu H, Cheng Y, Cheng Q, Li D. A hybrid intelligent method for three-dimensional short-term prediction of dissolved oxygen content in aquaculture. PLoS ONE 2018;13:e0192456.

[55] Fourati F, Chtourou M. A greenhouse control with feed-forward and recurrent neural networks. Simul Model Pract Theory 2007;15:1016–28. https://doi.org/10.1016/j.simpat.2007.06.001.

[56] Hongkang W, Li L, Yong W, Fanjia M, Haihua W, Sigrimis NA. Recurrent Neural Network Model for Prediction of Microclimate in Solar Greenhouse. IFAC-Pap 2018;51:790–5. https://doi.org/10.1016/j.ifacol.2018.08.099.

[57] Gong L, Yu M, Jiang S, Cutsuridis V, Pearson S. Deep Learning Based Prediction on Greenhouse Crop Yield Combined TCN and RNN. Sensors 2021;21:4537. https://doi.org/10.3390/s21134537.

[58] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212:372–85. https://doi.org/10.1016/j.apenergy.2017.12.051.

[59] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

[60] A. Ali, H.S. Hassanein, Time-Series Prediction for Sensing in Smart Greenhouses, in: GLOBECOM 2020 - 2020 IEEE Glob. Commun. Conf., IEEE, Taipei, Taiwan, 2020: pp. 1–6. https://doi.org/10.1109/GLOBECOM42002.2020.9322549.

[61] Hsieh H-C, Chiu Y-W, Lin Y-X, Yao M-H, Lee Y-J, Forecast LP, et al. Int. Conf. Pervasive Artif. Intell ICPAI, IEEE, Taipei, Taiwan 2020;2020:175–82. https://doi.org/10.1109/ICPAI51961.2020.00040.

[62] Jung D-H, Kim HS, Jhin C, Kim H-J, Park SH. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. Comput Electron

Agric 2020;173. https://doi.org/10.1016/j.compag.2020.105402 105402.

[63] Q. Zhao, Multi-point Prediction of Temperature and Humidity of Mushroom Based on CNN-GRU, Trans. Chin. Soc. Agric. Mach. (2020). https://doi.org/10.6041/j.issn.1000-1298.2020.09.033.

[64] Rumelhart DE, Hintont GE, Williams RJ. Learning representations by back-propagating errors 1986:4.

[65] Liguori A, Markovic R, Dam TTH, Frisch J, van Treeck C, Causone F. Indoor environment data time-series reconstruction using autoencoder neural networks. Build Environ 2021;191. https://doi.org/10.1016/j.buildenv.2021.107623 107623.

[66] Liu Y, Zhang Q, Song L, Chen Y. Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. Comput Electron Agric 2019;165. https://doi.org/10.1016/j.compag.2019.104964 104964.

[67] Kim EB, Park JH, Lee Y-S, Lim C. Two-dimensional attention-based multi-input LSTM for time series prediction, Commun. Stat. Appl. Methods 2021;28:39–57. Available from: https://doi.org/10.29220/CSAM.2021.28.1.039.

[68] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, G.W. Cottrell, A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction, in: Proc. Twenty-Sixth Int. Jt. Conf. Artif. Intell., International Joint Conferences on Artificial Intelligence Organization, Melbourne, Australia, 2017: pp. 2627–2633. https://doi.org/10.24963/ijcai.2017/366.

[69] Y. Yu, Y.-J. Kim, Two-Dimensional Attention-Based LSTM Model for Stock Index Prediction, J. Inf. Process. Syst. (n.d.) 12. https://doi.org/10.3745/JIPS.02.0121.

[70] H. Li, Y. Shen, Y. Zhu, Stock Price Prediction Using Attention-based Multi-Input LSTM, Dep. Comput. Sci. Eng. (n.d.) 16.

[71] Holzinger A, Malle B, Saranti A, Pfeifer B. Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI. Inf Fusion 2021;71:28–37. https://doi.org/10.1016/j.inffus.2021.01.008.

[72] Guo K, Hu Y, Qian Z, Liu H, Zhang K, Sun Y, et al. Optimized Graph Convolution Recurrent Neural Network for Traffic Prediction. IEEE Trans Intell Transp Syst 2021;22:1138–49. https://doi.org/10.1109/TITS.2019.2963722.

[73] Wang S-H, Govindaraj VV, Górriz JM, Zhang X, Zhang Y-D. Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. Int J Inf Fusion 2021;67:208–29. https://doi.org/10.1016/j.inffus.2020.10.004.

[74] La Gatta V, Moscato V, Postiglione M, Sperlí G. An Epidemiological Neural Network Exploiting Dynamic Graph Structured Data Applied to the COVID-19 Outbreak. IEEE Trans Big Data 2021;7:45–55. https://doi.org/10.1109/TBDATA.2020.3032755.

[75] A. Vyas, S. Bandyopadhyay, Semi-supervised Soil Moisture Prediction through Graph Neural Networks, ArXiv201203506 Cs. (2020). http://arxiv.org/abs/2012.03506 (accessed August 31, 2021).

[76] Park J. Physics-induced graph neural network: An application to wind-farm power estimation. Energy 2019;15.

[77] Chen R, Liu J, Wang F, Ren H, Zhen Z, Prediction G-B, et al. IEEE 3rd Stud. Conf. Electr. Mach. Syst SCEMS, IEEE, Jinan, China 2020;2020:982–7. https://doi.org/10.1109/SCEMS48876.2020.9352310.

[78] Jia-hui XU. Surface water quality prediction model based on graph neural network. J Zhejiang Univ Eng Sci 2021. https://doi.org/10.3785/j.issn.1008-973X.2021.04.001.

[79] Šmite D, Wohlin C, Gorschek T, Feldt R. Empirical evidence in global software engineering: a systematic review. Empir Softw Eng 2010;15:91–118. https://doi.org/10.1007/s10664-009-9123-y.

[80] Liu Y, Gong C, Yang L, Chen Y. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. Expert Syst Appl 2020;143. https://doi.org/10.1016/j.eswa.2019.113082 113082.