

Article

Comparison of Water Quality Prediction for Red Tilapia Aquaculture in an Outdoor Recirculation System Using Deep Learning and a Hybrid Model

Roongparit Jongjaraunsuk ¹, Wara Taparhudee ^{1,*}  and Pimlapat Suwannasing ²

¹ Department of Aquaculture, Faculty of Fisheries, Kasetsart University, Bangkok 10900, Thailand; ffsrrj@ku.ac.th

² Research Information Division, Kasetsart University Research and Development Institute (KURDI), Kasetsart University, Bangkok 10900, Thailand; rdiplp@ku.ac.th

* Correspondence: ffsiwrt@ku.ac.th

Abstract: In modern aquaculture, the focus is on optimizing production and minimizing environmental impact through the use of recirculating water systems, particularly in outdoor setups. In such systems, maintaining water quality is crucial for sustaining a healthy environment for aquatic life, and challenges arise from instrumentation limitations and delays in laboratory measurements that can impact aquatic animal production. This study aimed to predict key water quality parameters in an outdoor recirculation aquaculture system (RAS) for red tilapia aquaculture, including dissolved oxygen (DO), pH, total ammonia nitrogen (TAN), nitrite nitrogen (NO₂-N), and alkalinity (ALK). Initially, a random forest (RF) model was employed to identify significant factors for predicting each parameter, selecting the top three features from routinely measured parameters on the farm: DO, pH, water temperature (Temp), TAN, NO₂-N, and transparency (Trans). This approach aimed to streamline the analysis by reducing variables and computation time. The selected parameters were then used for prediction, comparing the performance of convolutional neural network (CNN), long short-term memory (LSTM), and CNN-LSTM models across different epochs (1000, 3000, and 5000). The results indicated that the CNN-LSTM model at 5000 epochs was effective in predicting DO, TAN, NO₂-N, and ALK, with high R² values (0.815, 0.826, 0.831, and 0.780, respectively). However, pH prediction showed lower efficiency with an R² value of 0.377.

Keywords: red tilapia; outdoor recirculating aquaculture system; water quality prediction; machine learning models



Citation: Jongjaraunsuk, R.; Taparhudee, W.; Suwannasing, P. Comparison of Water Quality Prediction for Red Tilapia Aquaculture in an Outdoor Recirculation System Using Deep Learning and a Hybrid Model. *Water* **2024**, *16*, 907. <https://doi.org/10.3390/w16060907>

Academic Editor: Stefano Amalfitano

Received: 13 February 2024

Revised: 15 March 2024

Accepted: 18 March 2024

Published: 21 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nile tilapia (*Oreochromis niloticus* Linn.) is an important freshwater fish that is widely cultivated around the world [1]. Currently, farmers prioritize maximizing productivity within confined spaces while minimizing environmental impact. To achieve this, recirculating aquaculture systems (RASs) have been widely implemented in Europe and the USA, especially indoors [2]. Maintaining water quality is pivotal in these systems [3–7], necessitating the continuous monitoring of parameters such as dissolved oxygen (DO) levels, water temperature (Temp), acidity, alkalinity (ALK), water clarity, total ammonia nitrogen (TAN), and nitrite nitrogen (NO₂-N). Measurements involve field equipment and laboratory assessments, often requiring chemicals, instruments, labor, time, and expenses [8]. In open recirculating systems, water quality parameters may swiftly fluctuate due to environmental changes, thus demanding more resources than closed systems. Additionally, equipment breakdowns and delayed laboratory measurements can pose challenges, potentially impacting aquatic animal rearing.

However, to date, most research has used water quality relationships to predict outcomes, often assuming linear correlations, which might not always yield accurate

predictions, due to nonlinear relationships and environmental variations. In recent years, machine learning (ML) is a technique that has become an increasingly important tool in data analysis, being applied in various fields of research, including applications related to water quality, or work in aquaculture. For example, Palani et al. [9] employed artificial neural networks (ANNs) for marine water quality, while Castrillo and García [10] applied multi-factor linear regression (MLR) and random forest (RF) models for predicting river water quality. Zambrano et al. [8] utilized RF, MLR models, and an ANN for predicting water quality in fish farming reservoirs. Anand et al. [11] and Ye et al. [12] opted for CNNs in their models. Additionally, Hu et al. [13] used a deep long short-term memory (LSTM) network for cage-cultured environments, and Liu et al. [14] employed LSTM deep neural networks in an internet of things (IoT) setting for predicting water quality in cage-cultured environments. Ahmed et al. [15] applied gradient boosting and the multi-layer perceptron (MLP). Juna et al. [16] proposed a nine-layer MLP model with k-nearest neighbors (KNN) imputation. Li et al. [17] recommended support vector machine (SVM) for industrial aquaculture. Wang et al. [18] suggested SVM for dissolved oxygen, MLP for nonlinear modeling, and LSTM for dynamic patterns in precise water quality prediction. This diversification of ML techniques across various studies underscores the evolving and versatile nature of ML applications in understanding and forecasting water quality.

In recent advancements, various models have been integrated to enhance prediction accuracy [19]. Notably, da Silva et al. [20] introduced a novel toxicity-warning sensor with the linear concentration addition (LCA) model and ML for water quality monitoring. Chen et al. [21] proposed intelligent variable-flow technology for improved water quality control. Meanwhile, Yang et al. [22] utilized a hybrid deep learning approach (CNN, gated recurrent unit (GRU), and attention mechanism) for RAS water quality prediction. Additionally, Wu and Wang [23] introduced the artificial neural network-wavelet transform-long short-term memory (ANN-WT-LSTM) model for Jinjiang River, surpassing others in water quality prediction. Zhou et al. [24] presented the wavelet-autoregressive integrated moving average-gated recurrent unit (W-ARIMA-GRU) model for Beijing's water, combining wavelet decomposition, ARIMA, and GRU. Chen et al. [25] developed LSTM and attention-based long short-term memory (AT-LSTM) models for forecasting Australia's Burnett River water quality. Cai et al. [26] proposed a Kalman-filter-aided LSTM with attention for improved accuracy on Haimen Bay data. Farzana et al. [27] used XGBoost and GRU models for Toowoomba reservoirs, providing insights for climate-aware water management. These studies widely attest to the efficacy and success of this methodology.

However, hybrid models often involve a relatively high number of features, leading to complex processing and potentially lengthy computations. Therefore, the objective of the current study was to predict essential water quality parameters in large-scale, open recirculating red tilapia systems. Notably, there is limited research on refining the predictive model by reducing the number of features. This was achieved through the application of an RF model, complemented by the utilization of the CNN, LSTM, and CNN-LSTM models. Adjustments in epochs (1000, 3000, and 5000) were made to optimize accuracy levels. The findings should help to advance the development of outdoor RAS methodologies.

2. Materials and Methods

2.1. Farming System and Data Collection

The data were collected from a red tilapia farm in Buriram province, northeast Thailand (15°04'01.9" N 102°47'20.3" E). The farm used an RAS consisting of 3 treatment ponds for the water inlet, 4 nursing ponds (each 1600 m²), 18 grow-out ponds (each 1600 m²), and 5 treatment ponds for the water outlet. All the nursing and grow-out ponds were lined with polyethylene. Three grow-out ponds were selected as experimental ponds (1, 2, and 3), as shown in Figure 1. All the experimental ponds were under the same management regime, involving aeration using four 3 horsepower (Hp) aerators that operated continuously (Figure 2a). The fish were fed with 35% protein pelleted feed 3 times a day (08.00, 11.30, and 16.30) using an automatic feeder (Figure 2b). The average starting weight of fish raised

was about 200 g. The stocking density was 19,000 fish/pond (about 12 fish/m²). The fish weight was assessed manually and randomly twice a month using about 45 fish/pond after anesthetization with clove oil (10 µL/L). The average fish weight on the day of harvest was about 1000 g. The average survival rate was 95%. The rearing period was approximately 90 days.

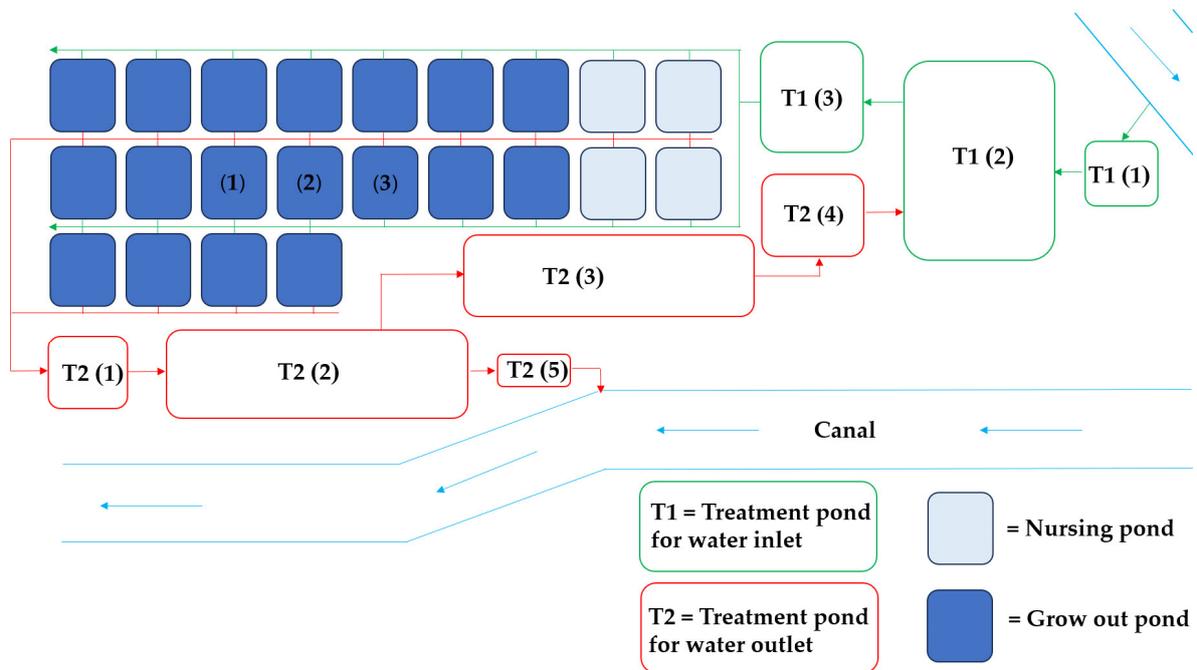


Figure 1. Layout of fish farm; T1 = treatment ponds for water inlet with the areas of T1 (1) = 1600 m², T1 (2) = 9200 m², and T1 (3) = 4600 m²; T2 = treatment ponds for water outlet with the areas of T2 (1) = 2000 m², T2 (2) = 8200 m², T2 (3) = 7200 m², T2 (4) = 3600 m², and T2 (5) = 800 m². Note: Green arrows show water inlet flow direction, red arrows show water outlet flow direction, and blue arrows show the direction of water flow in canal.

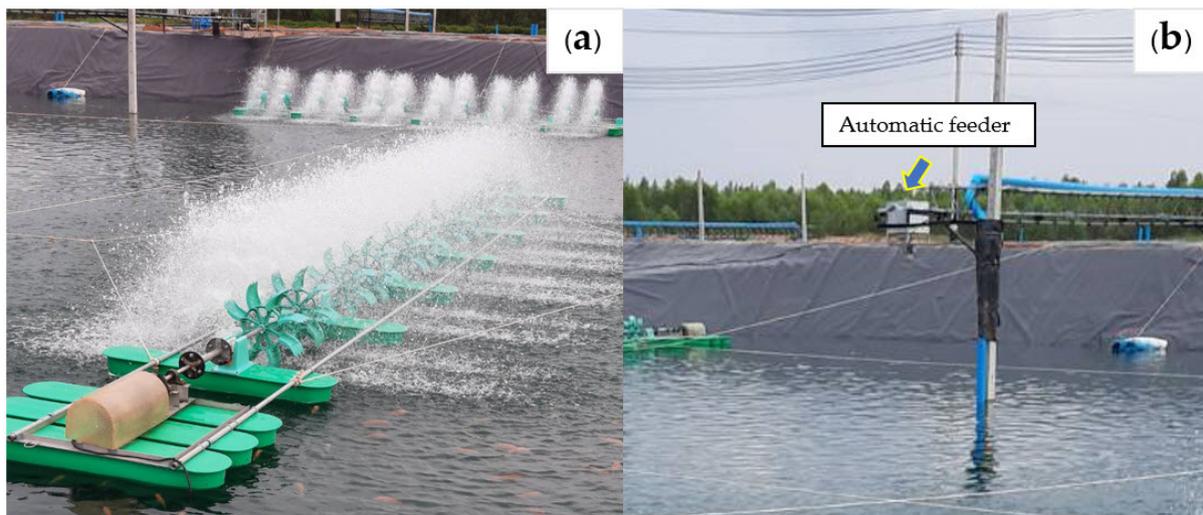


Figure 2. Important equipment used in ponds: (a) 3 Hp aerators that operated continuously and (b) automatic feeder.

2.2. Water Quality Measurement

A total of 2250 water samples were collected for analysis of DO, Temp, pH, TAN, NO₂-N, ALK, and Trans. The DO and Temp were monitored using a YSI Pro20i instrument

(YSI; Yellow Springs, OH, USA). The pH was measured using a YSI pH100A instrument (YSI; USA) and Trans was measured using a Secchi disc. The TAN, NO₂-N, and ALK were sampled for analysis in the laboratory according to the method of APHA [28]. All parameters were monitored every day in the morning between 07.00 and 08.00 throughout the 4 months of the growth cycle.

2.3. Pre-Processing Dataset

Before running the processes, data cleaning was performed by checking for missing data, which was addressed by either removing entries with missing values or imputing them using statistical methods (mean, median, or mode). Approximately 0.2% of the total data were affected. Additional cleaning involved correcting any incorrect data or formatting issues, such as pH values of 15 and 7.5, which impacted approximately 0.5% of the total data.

2.4. Feature Selection

RF was applied to identify important features for predicting each water quality parameter. The process of selecting important features using RandomForestRegressor entails utilizing the model's built-in feature importance attribute. During training, feature importance was calculated by measuring how each feature decreases impurity across decision trees, highlighting their contributions to predictive performance. All the processes are shown in Table 1. The training performances of the RF were assessed by computing the mean absolute error (MAE). These metrics provide a numerical measure of how well the model captures the real-world conditions. After the selection process, the focus narrowed down to processing only the top 3 features. This adjustment aimed to expedite processing and simplify the task, reducing both processing time and complexity.

Table 1. The key steps of feature selection using RF.

Key Step	Process
Import libraries	Pandas for data manipulation. RandomForestRegressor for building the regression model. Other libraries for data processing, evaluation, and visualization.
Load and preprocess data	Load a csv dataset and select relevant features and the target variable.
Train–test split	Split the data into training and testing sets.
Initialize and train a RandomForestRegressor with specific parameters	Initialize and train a RandomForestRegressor with specific parameters. The code configures the regressor with 100 trees, a random seed of 42 for consistency, a maximum tree depth of 10, and a maximum of 10 leaf nodes per tree. Then, it trains the regressor using the given dataset.
Model evaluation	Evaluate the model on both training and testing sets using MAE.
Visualize predictions	Create a scatter plot to visualize predicted vs. actual values.
Feature importance bar graph	Calculate and display a bar graph showing the importance of each feature in predicting the parameter.

2.5. Data Processing, Analysis, and Visualization

Python (version 3.9), in a colab notebook setting, was used for the essential tasks encompassing deep learning and data analysis. Data processing: the python library (pandas) was utilized for data manipulation, while scikit-learn was used for dataset splitting and feature scaling (min–max scaler), and TensorFlow was used for handling neural network data structures. The analysis used TensorFlow's Keras components to construct neural network models, such as sequential, conv1D, maxpooling1D, LSTM, dense, flatten, and dropout alongside importing evaluation metrics; root mean square error (RMSE), MAE, normalized root mean square error (NRMSE), nash-sutcliffe efficiency (NSE), and the coefficient of determination (R^2) from scikit-learn to assess model performance. Matplotlib (version 3.8)

was used for data visualization, enabling the creation of graphs, charts, and other visual representations of the data and model outputs within the colab notebook.

Initially, the data were loaded and divided into a training set (80%) and a testing set (20%). Next, 3 models (CNN, LSTM, and a hybrid CNN–LSTM) were used for analysis. All the models underwent fine-tuning by progressively increasing the number of epochs (1000, 3000, and 5000). Here, epochs are the number of times a model goes through the entire training dataset during training. The model structures are shown in Table 2.

Table 2. Structures of developed models.

Model	Structure
CNN	<pre> Model = sequential () Model.add (conv1D (1024, kernel_size = 3, activation = 'relu', Input_shape = (x_train.shape [1], 1))) Model.add (maxpooling1D (pool_size =1)) Model.add (flatten ()) Model.add (dense (128, activation = 'relu')) Model.add (dropout (0.5)) Model.add (dense (1, activation = 'linear')) Model.compile (optimizer = 'adam', loss = 'mean_squared_error', metrics = ['mae']) Calculate evaluation metrics: MAE, RMSE, NRMSE, NSE, and R² </pre>
LSTM	<pre> Model = sequential () Model.add (LSTM (300, return_sequences = True)) Model.add (LSTM (300)) Model.add (dense (128, activation = 'relu')) Model.add (dropout (0.5)) Model.add (dense (1, activation = 'linear')) Model.compile (optimizer = 'adam', loss = 'mean_squared_error', metrics = ['mae']) Calculate evaluation metrics: MAE, RMSE, NRMSE, NSE, and R² </pre>
CNN-LSTM	<pre> Model = sequential () Model.add (conv1D (1024, kernel_size = 3, activation = 'relu', Input_shape = (x_train.shape [1], 1))) Model.add (maxpooling1D (pool_size =1)) Model.add (LSTM (300, return_sequences = True)) Model.add (LSTM (300)) Model.add (dense (128, activation = 'relu')) Model.add (dropout (0.5)) Model.add (dense (1, activation = 'linear')) Model.compile (optimizer = 'adam', loss = 'mean_squared_error', metrics = ['mae']) Calculate evaluation metrics: MAE, RMSE, NRMSE, NSE, and R² </pre>

2.6. Performance Metrics

The performance of the models was assessed using 5 metrics: RMSE, MAE, NRMSE, NSE, and R². These metrics were calculated using Equations (1)–(5).

$$\text{RMSE} = \sqrt{\frac{\sum_{N}^{i=1} (y_{\text{measured},i} - y_{\text{predict},i})^2}{N}} \quad (1)$$

$$\text{MAE} = \frac{\sum_{N}^{i=1} |y_{\text{measured},i} - y_{\text{predict},i}|}{N} \quad (2)$$

$$\text{NRMSE} = \sqrt{\frac{\sum_{N}^{i=1} (y_{\text{measured},i} - y_{\text{predict},i})^2}{y_{\text{max}} - y_{\text{min}}}} \quad (3)$$

where y_{measured} is the observed values, y_{predict} is the predicted values, N is the total number of variables, y_{max} is the maximum value, and y_{min} is the minimum value.

$$\text{NSE} = 1 - \frac{\sum_{t=1}^N (\text{Obs}_t - \text{Sim}_t)^2}{\sum_{t=1}^N (\text{Obs}_t - \bar{\text{Obs}})^2} \quad (4)$$

where Obs_t is the observed value at time t , Sim_t is the simulated (predicted) value at time t , and $\bar{\text{Obs}}$ is the mean of the observed values.

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}} \quad (5)$$

where SS_{res} is the sum of squares of residuals (also known as the sum of squared errors or SSE), which represents the difference between the predicted values and the actual values; and SS_{tot} is the total sum of squares, which measures the total variance of the dependent variable (the target) from its mean. Moreover, the calculation times of each model were measured.

2.7. Ethical Statement

The study protocol for fish care and experiments was reviewed and approved by the Kasetsart University institutional animal care and use committee (ACKU 66-FIS-004). This study followed Arrive guidelines (<https://arriveguidelines.org>, accessed on 19 August 2023). All methods were performed in accordance with the relevant guidelines and regulations.

3. Results

3.1. Water Quality

The average values for DO, Temp, pH, TAN, $\text{NO}_2\text{-N}$, ALK, and Trans were 6.15 ± 1.02 mg/L, 22.25 ± 0.87 °C, 7.35 ± 0.05 , 0.81 ± 0.41 mg/L, 0.78 ± 0.13 mg/L, 72.43 ± 9.30 mg/L, and 34.90 ± 6.90 cm, respectively. The average final fish weight, average daily growth (ADG), and survival rate were 834.17 ± 102.35 g, 5.94 ± 1.02 g/day, and $97.37 \pm 2.71\%$, respectively, as shown in Table 3.

Table 3. Mean \pm standard deviation and standard quality range of dataset variables.

Parameters	Value	Standard Quality	Reference
Culture details			
Week of culture (WOC)	14		
Initial fish weight (g/fish)	254.67 ± 7.09		
Final weight (g/fish)	834.17 ± 102.35		
ADG (g/fish/day)	5.94 ± 1.02		
Survival rate (%)	97.37 ± 2.71		
Water quality parameter			
DO (mg/L)	6.15 ± 1.02	>3	[29,30]
Temp (°C)	22.25 ± 0.87	25–32	[31]
pH	7.35 ± 0.05	7–8	[32,33]
TAN (mg/L)	0.81 ± 0.41	<0.5	[34]
$\text{NO}_2\text{-N}$ (mg/L)	0.78 ± 0.13	<0.5	[32,35]
ALK (mg/L)	72.43 ± 9.30	75–400	[32,36]
Trans (cm)	34.90 ± 6.90	15–40	[37,38]

3.2. Important Features for Each Water Quality Parameter Prediction

The MAE values for the training performances of the RF model were as follows: DO (0.247), pH (0.053), TAN (0.246), $\text{NO}_2\text{-N}$ (0.093), and ALK (2.127). These results are visually presented in Figure 3.

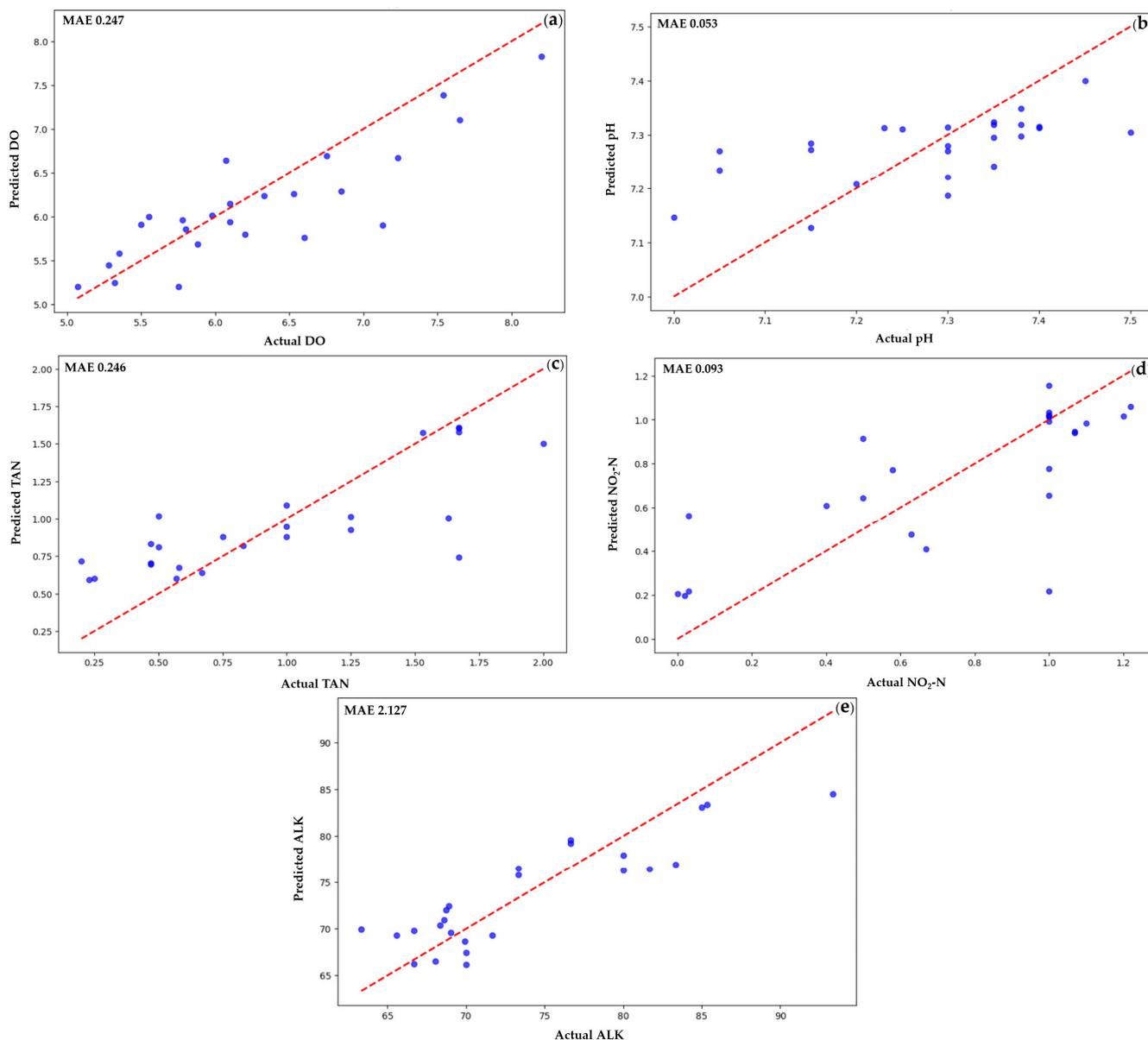


Figure 3. Prediction performances of RF by MAE for (a) DO, (b) pH, (c) TAN, (d) NO₂-N, and (e) ALK. The red line in the scatter plot represents the MAE line, which is a measure of how different the predicted values are from the actual values. The blue dots in the scatter plot represent the actual value measurements. The horizontal position of each dot shows the actual value, and the vertical position shows the residual value, which is the difference between the predicted value and the actual value. In linear regression, the goal is to fit a line through the data points in a way that minimizes the residuals. The MAE line can be used to assess how well the fitted line meets this goal. A lower MAE value indicates that the predictions are, on average, closer to the actual values.

The top three most influential features for each water quality parameter were selected using RF. Figure 4a outlines feature importance for predicting DO in a water quality model, highlighting Temp (0.751) as the most critical, followed by NO₂-N (0.065) and ALK (0.053). In Figure 4b, ALK (0.247) is crucial for pH prediction, followed by DO (0.241) and Trans (0.195). Figure 4c indicates the role of Trans (0.237) in TAN production prediction, followed by ALK (0.218) and DO (0.180). Figure 4d highlights the dominance of ALK (0.372) in NO₂-N prediction, followed by Temp (0.255) and TAN (0.145). Lastly, Figure 4e shows the importance of Trans (0.391) in ALK prediction, followed by TAN (0.188) and NO₂-N (0.184).

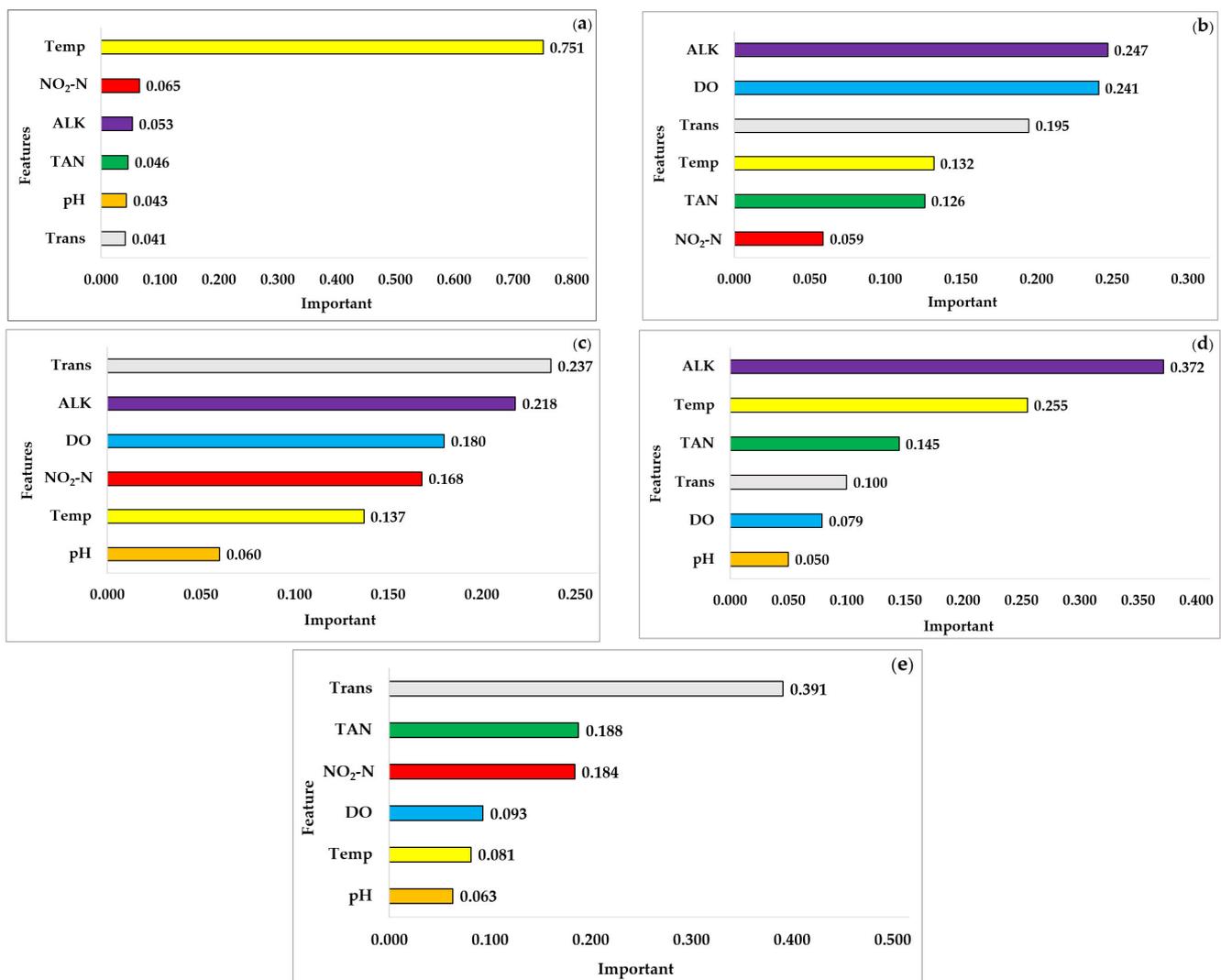


Figure 4. Important features for prediction of (a) DO, (b) pH, (c) TAN, (d) NO₂-N, and (e) ALK.

3.3. Predictive Efficiency

The performance metrics (RMSE, MAE, NRMSE, NSE, and R^2) were compared among the three models across various epochs (1000–5000). Notably, the CNN-LSTM model, particularly at 5000 epochs, exhibited superior predictive capabilities for key water quality parameters of DO, pH, TAN, NO₂-N, and ALK. This model consistently demonstrated lower RMSE, MAE, and NRMSE values compared to the other models. Furthermore, NSE values were consistently higher than those of the other models. Specifically, the R^2 values for the CNN-LSTM model at 5000 epochs reached peaks at 0.815 (DO), 0.826 (TAN), 0.831 (NO₂-N), and 0.780 (ALK). However, the pH prediction notably underperformed, with an R^2 of only 0.377. Additionally, the calculation times for this model were approximately 15 min, as shown in Table 4. Following the application of the developed model, graphs displaying both observed and predicted values for each crucial water quality parameter (DO, pH, TAN, NO₂-N, and ALK) are depicted in Figures 5–9.

Table 4. Performance comparison of different models in each epoch for predicting DO, pH, TAN, NO₂-N, and ALK, where bold indicates best performing model for each water parameter.

Parameter	Model	RMSE	MAE	NRMSE	NSE	R ²	Time
DO	CNN 1000 epoch	0.396	0.312	0.100	0.696	0.755	0 min 28 s
	CNN 3000 epoch	0.394	0.291	0.099	0.698	0.759	1 min 25 s
	CNN 5000 epoch	0.396	0.301	0.100	0.764	0.756	2 min 25 s
	LSTM 1000 epoch	0.455	0.356	0.114	0.639	0.677	3 min 29 s
	LSTM 3000 epoch	0.442	0.335	0.111	0.733	0.695	9 min 29 s
	LSTM 5000 epoch	0.448	0.349	0.112	0.778	0.688	16 min 30 s
	CNN-LSTM 1000 epoch	0.486	0.400	0.122	0.708	0.632	3 min 27 s
	CNN-LSTM 3000 epoch	0.386	0.300	0.097	0.784	0.768	9 min 23 s
	CNN-LSTM 5000 epoch	0.344	0.240	0.086	0.836	0.815	15 min 55 s
pH	CNN 1000 epoch	0.421	0.407	0.795	−2.492	−11.577	0 min 28 s
	CNN 3000 epoch	0.137	0.112	0.259	−0.513	−0.340	1 min 26 s
	CNN 5000 epoch	0.114	0.092	0.215	0.116	0.080	2 min 21 s
	LSTM 1000 epoch	0.138	0.112	0.261	−3.978	−0.355	3 min 58 s
	LSTM 3000 epoch	0.110	0.088	0.207	−0.524	0.148	9 min 30 s
	LSTM 5000 epoch	0.134	0.108	0.252	0.188	−0.269	13 min 21 s
	CNN-LSTM 1000 epoch	0.113	0.092	0.214	−2.306	0.088	3 min 26 s
	CNN-LSTM 3000 epoch	0.128	0.102	0.242	−1.275	−0.165	9 min 16 s
	CNN-LSTM 5000 epoch	0.094	0.075	0.177	0.477	0.377	15 min 35 s
TAN	CNN 1000 epoch	0.361	0.271	0.101	0.663	0.651	0 min 30 s
	CNN 3000 epoch	0.283	0.207	0.079	0.792	0.786	1 min 26 s
	CNN 5000 epoch	0.267	0.192	0.075	0.793	0.808	2 min 24 s
	LSTM 1000 epoch	0.445	0.325	0.125	0.694	0.468	3 min 28 s
	LSTM 3000 epoch	0.319	0.223	0.089	0.820	0.727	8 min 28 s
	LSTM 5000 epoch	0.335	0.223	0.094	0.846	0.700	13 min 28 s
	CNN-LSTM 1000 epoch	0.299	0.237	0.084	0.713	0.760	3 min 28 s
	CNN-LSTM 3000 epoch	0.324	0.229	0.091	0.826	0.719	9 min 20 s
	CNN-LSTM 5000 epoch	0.255	0.156	0.071	0.895	0.826	15 min 28 s
NO ₂ -N	CNN 1000 epoch	0.173	0.122	0.104	0.764	0.772	0 min 29 s
	CNN 3000 epoch	0.174	0.109	0.104	0.772	0.771	1 min 29 s
	CNN 5000 epoch	0.193	0.130	0.116	0.802	0.717	2 min 32 s
	LSTM 1000 epoch	0.259	0.198	0.155	0.737	0.491	2 min 44 s
	LSTM 3000 epoch	0.202	0.143	0.121	0.729	0.690	8 min 41 s
	LSTM 5000 epoch	0.186	0.130	0.112	0.760	0.736	14 min 33 s
	CNN-LSTM 1000 epoch	0.176	0.123	0.106	0.778	0.764	3 min 20 s
	CNN-LSTM 3000 epoch	0.155	0.089	0.093	0.807	0.817	9 min 20 s
	CNN-LSTM 5000 epoch	0.149	0.078	0.089	0.814	0.831	15 min 31 s
ALK	CNN 1000 epoch	5.993	4.238	0.150	0.430	0.310	0 min 33 s
	CNN 3000 epoch	5.100	3.482	0.127	0.507	0.500	1 min 26 s
	CNN 5000 epoch	4.134	2.979	0.103	0.578	0.672	2 min 24 s
	LSTM 1000 epoch	6.785	4.454	0.170	0.228	0.115	3 min 28 s
	LSTM 3000 epoch	6.318	4.172	0.158	0.637	0.233	8 min 32 s
	LSTM 5000 epoch	6.554	4.592	0.164	0.715	0.174	14 min 27 s
	CNN-LSTM 1000 epoch	7.613	6.171	0.190	0.529	−0.114	3 min 25 s
	CNN-LSTM 3000 epoch	4.701	3.653	0.118	0.684	0.575	9 min 29 s
	CNN-LSTM 5000 epoch	3.384	2.524	0.085	0.739	0.780	15 min 27 s

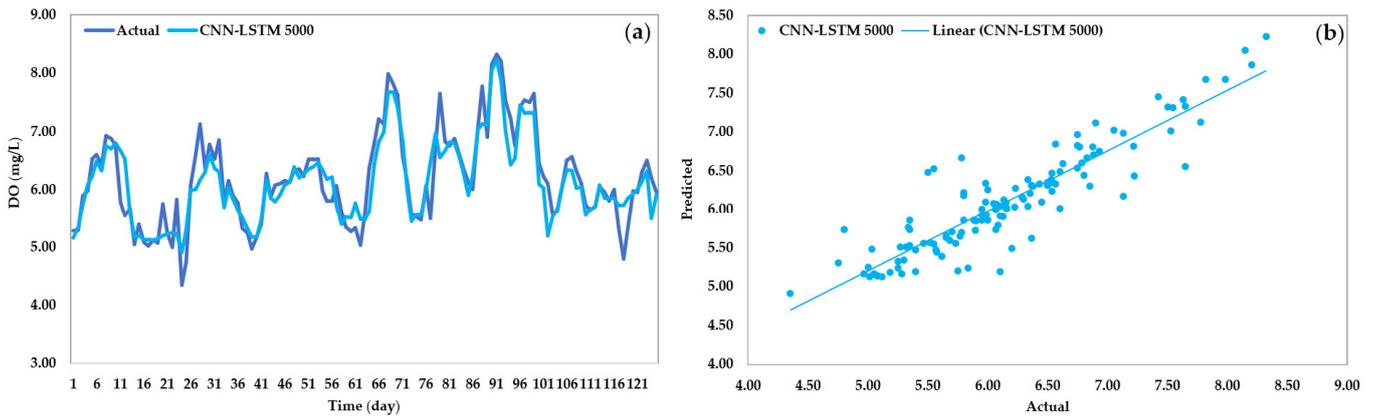


Figure 5. Actual and predicted values (a), along with a scatter graph (b), obtained from the CNN-LSTM model after 5000 epochs of DO.

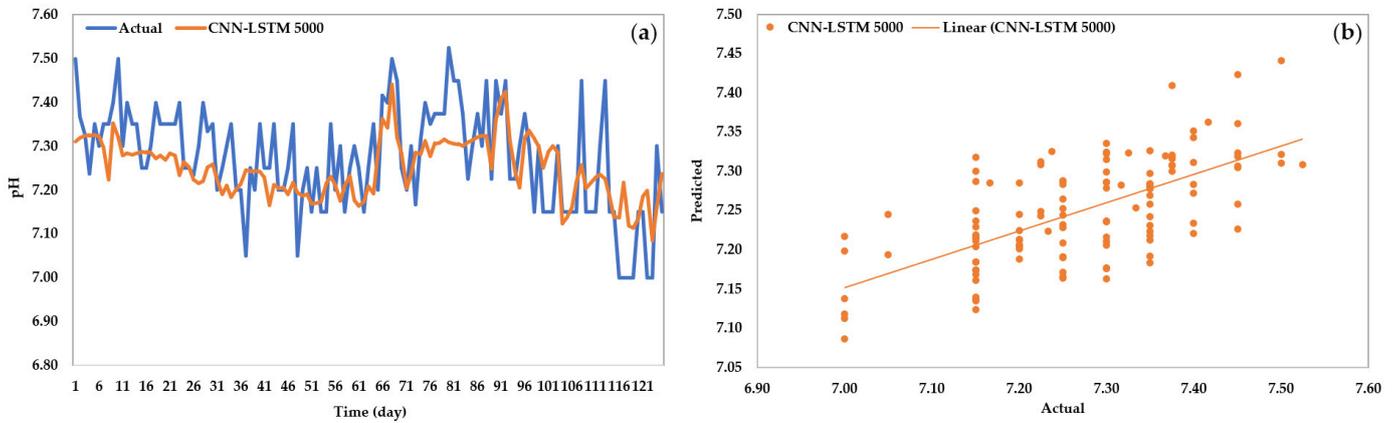


Figure 6. Actual and predicted values (a), along with a scatter graph (b), obtained from the CNN-LSTM model after 5000 epochs of pH.

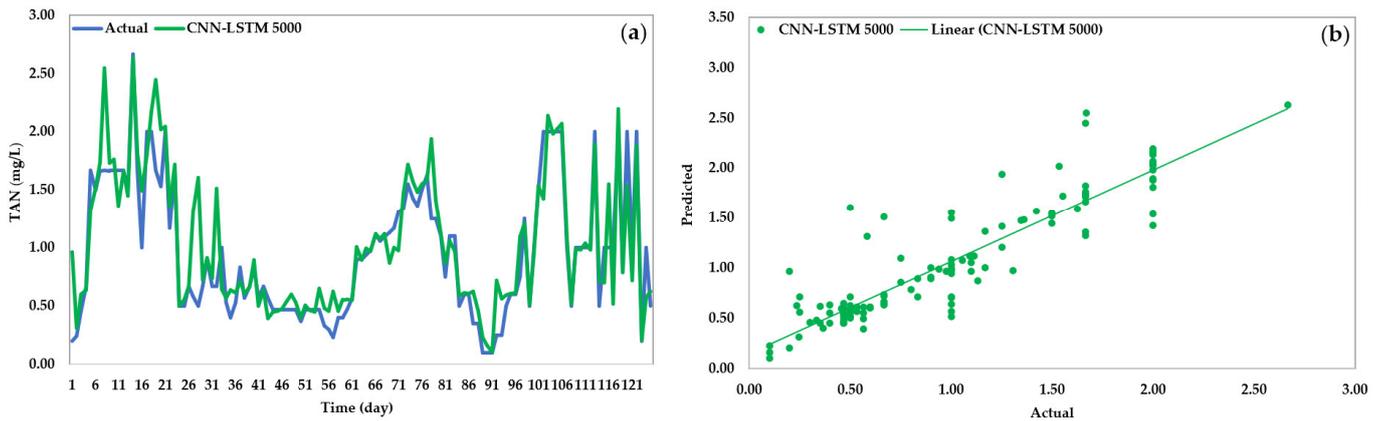


Figure 7. Actual and predicted values (a), along with a scatter graph (b), obtained from the CNN-LSTM model after 5000 epochs of TAN.

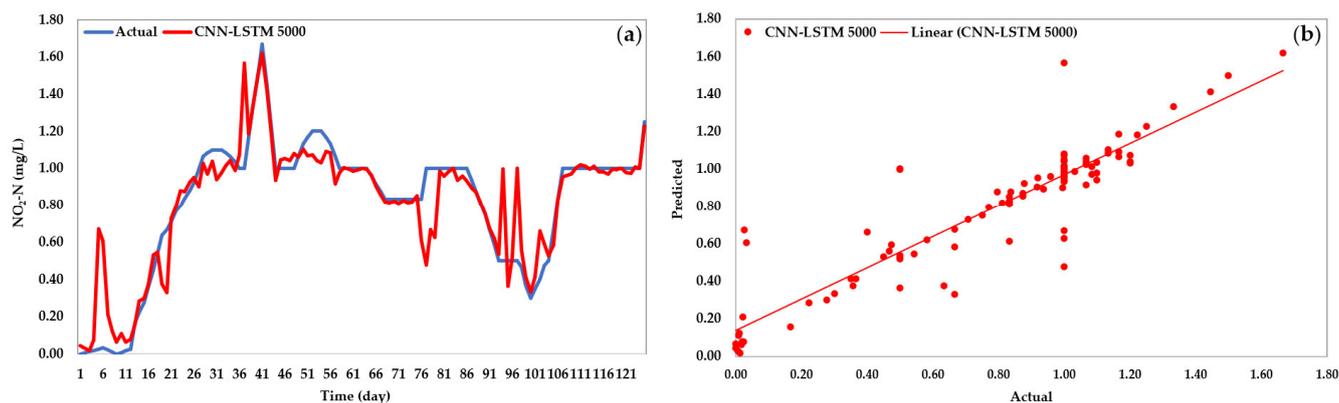


Figure 8. Actual and predicted values (a), along with a scatter graph (b), obtained from the CNN-LSTM model after 5000 epochs of NO₂-N.

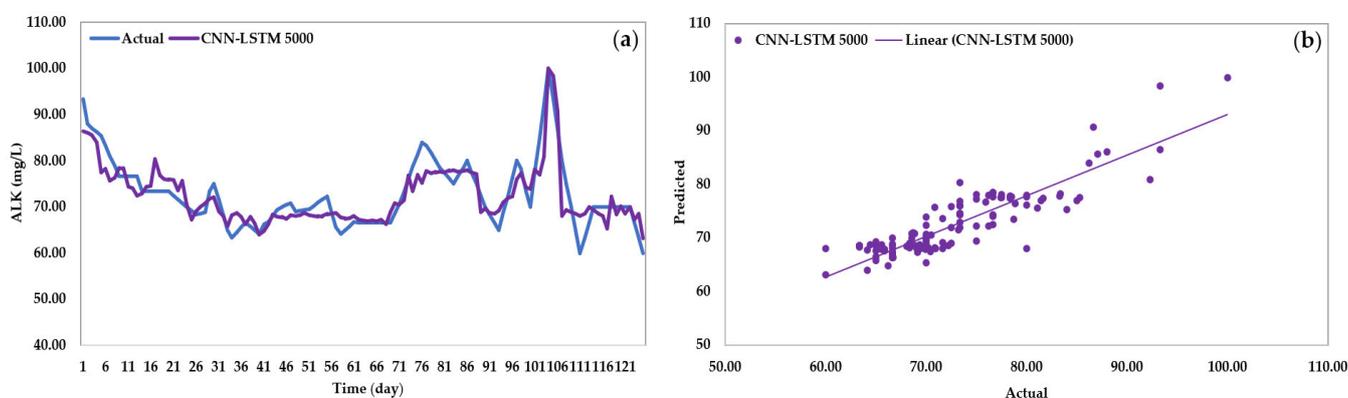


Figure 9. Actual and predicted values (a), along with a scatter graph (b), obtained from the CNN-LSTM model after 5000 epochs of ALK.

4. Discussion

The CNN and LSTM model combination was clearly the best for prediction tasks due to their complementary strengths. CNNs excel in extracting spatial features, making them well suited for data like images with spatial patterns, while LSTMs specialize in capturing temporal dependencies, fitting perfectly for sequential data such as time series or text [39–41]. This coupling allows for hierarchical feature learning, where the CNNs extract features and the LSTMs sequentially process them for deeper insight [42,43]. Furthermore, this model exhibits acceptable computation times for predictions when compared to laboratory analysis.

The findings of this study are consistent with prior research that utilized a hybrid CNN-LSTM for predicting water quality. In 2020, Baek et al. [44] demonstrated the accuracy of a CNN-LSTM model in simulating water quality in the Nakdong River basin, achieving ‘very good’ performance and proving valuable for precise water level and quality simulation. Li et al. [45] employed a CNN-LSTM model to compute runoff in the Elbe River basin, Germany, using two-dimensional rainfall radar maps. This model proved beneficial for assessing water availability and providing flood alerts in river basin management. Additionally, in 2023, Li et al. [46] introduced CLATT, a CNN-LSTM-attention model, enhancing wastewater quality prediction accuracy with a sliding window method.

The advantages of the above hybrid model combined with the selected features conducted using RF include that our results can be explained as temperature directly impacting oxygen solubility in the water, with warmer temperatures decreasing oxygen levels [47], while other parameters may have a lesser effect.

In forecasting TAN levels, Trans is used to measure clarity and particle presence, indicating organic matter [47]. This may relate to TAN levels, as organic content influences ammonia levels [48]. ALK also plays a role by influencing pH [49], where higher pH

levels can elevate TAN production and toxicity [50]. In addition, the presence of nitrifying bacteria, reliant on DO, affects the efficiency of nitrification, consequently lowering TAN levels with higher DO concentrations [51].

For $\text{NO}_2\text{-N}$ prediction, in the nitrogen cycle, it is not solely the ALK but also the presence of oxygen and specialized bacteria that govern the process; however, these relations are quite complex. ALK can significantly influence the solubility of $\text{NO}_2\text{-N}$ in water, as a higher ALK level may indirectly impact the nitrogen cycle and alter the speciation of nitrogen compounds [52]. Additionally, ALK facilitates the volatilization of ammonia. In turn, temperature affects the production and consumption rate of $\text{NO}_2\text{-N}$, as it is generated by ammonia oxidation—a process catalyzed by temperature [53]. Furthermore, TAN acts as a precursor to $\text{NO}_2\text{-N}$, as bacteria-mediated ammonia oxidation leads to the formation of $\text{NO}_2\text{-N}$ [47].

With ALK prediction, Trans has been reported as a pivotal factor, since Trans serves as an indicator of water clarity, influenced by suspended particles such as algae and sediment [47]. These particles absorb sunlight, potentially reducing available light for photosynthesis, subsequently impacting the growth of aquatic plants and algae, which, in turn, affects ALK levels [54]. TAN influences ALK by interacting with bicarbonate ions, leading to a reduction in alkalinity levels, with elevated TAN concentrations contributing to this reduction. Furthermore, specific waterborne bacteria, such as *Nitrobacter* and *Nitrospira*, can convert $\text{NO}_2\text{-N}$ into nitrate-nitrogen ($\text{NO}_3\text{-N}$), a process that releases hydrogen ions (H^+). This conversion indirectly leads to an increase in ALK by shifting the pH towards a more neutral or slightly alkaline state [47]. Trans itself does not directly affect ALK; instead, it refers to the clarity or clearness of water. ALK, on the other hand, is a measure of the water's ability to resist changes in pH. However, the factors influencing Trans, such as suspended particles or dissolved substances, can indirectly influence ALK. These factors have the potential to absorb or adsorb alkaline substances.

Predicting pH directly from the data collected might be difficult because there are factors like CO_2 , minerals, pollution, and biological activities that can affect pH [55], but we did not include them in this study. So, it is not practical to predict pH only based on our available data. Also, these factors affect other parameters, not just pH.

However, notably, the effectiveness of any model combination depends heavily on the nature of the data and the specific problem at hand. While CNN–LSTM hybrids have shown promise in certain applications, other architectures or models might perform better in different scenarios. The choice of model often involves empirical testing and experimentation to find the most suitable one for a particular task.

5. Conclusions

The red tilapia outdoor recirculating study focused on predicting crucial water quality indicators DO, pH, TAN, $\text{NO}_2\text{-N}$, and ALK. Three key features were selected using the RF model; different models (CNN, LSTM, and a hybrid CNN–LSTM) were tested by varying the epochs from 1000 to 5000. Using the CNN–LSTM model at 5000 epochs demonstrated notably high-performance metrics (RMSE, MAE, NRMSE, and R^2) for the parameters DO, TAN, $\text{NO}_2\text{-N}$, and ALK. However, the prediction of pH had comparatively lower results. These outcomes might have been influenced by external environmental factors. A limitation of this study is the absence of measurements for other environmental parameters, such as meteorological data. This is crucial due to environmental variations and seasonal effects that can significantly impact water quality in outdoor settings, along with other biological data.

Author Contributions: Conceptualization, R.J. and W.T.; methodology, R.J., W.T. and P.S.; software, R.J. and W.T.; validation, R.J., W.T. and P.S.; formal analysis, R.J. and W.T.; investigation, R.J. and W.T.; resources, R.J.; data curation, R.J., W.T. and P.S.; writing—original draft preparation, R.J. and W.T.; writing—review and editing, R.J. and W.T.; visualization, R.J. and W.T.; supervision, W.T.; project administration, R.J.; funding acquisition, R.J. and W.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Faculty of Fisheries, Kasetsart University, under the project titles ‘A decision support system using machine learning techniques for finding best practices of red tilapia (*Oreochromis Niloticus* Linn.) reared in an outdoor recirculating aquaculture system’. For financing the study, and the research facilities were made available by Patthamarach farm in Lam Plai Mat District, Buriram province, Thailand.

Data Availability Statement: The corresponding author can provide the datasets for this work upon reasonable request.

Acknowledgments: We are thankful to the staff of our research project and aquacultural engineering laboratory, Department of Aquaculture, Faculty of Fisheries, Kasetsart University, for their support during the trials.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Amin, M.; Musdalifah, L.; Ali, M. Growth performances of Nile Tilapia, *Oreochromis niloticus*, reared in recirculating aquaculture and active suspension systems. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *441*, 012135. [\[CrossRef\]](#)
- Dalsgaard, J.; Lund, I.; Thorarinsdottir, R.; Drengstig, A.; Arvonen, K. Farming different species in RAS in NORDIC countries: Current status and future perspectives. *Aquac. Eng.* **2002**, *53*, 2–13. [\[CrossRef\]](#)
- El-Sayed, A.F.M. Effect of stocking density and feeding levels on growth and feed efficiency of Nile tilapia (*Oreochromis niloticus* L.) fry. *Aquac. Res.* **2022**, *33*, 621–626. [\[CrossRef\]](#)
- Gibtan, A.; Getahun, A.; Mengistou, S. Effect of stocking density on the growth performance and yield of Nile tilapia (*Oreochromis niloticus* L., 1758) in a cage culture system in Lake Kuriftu, Ethiopia. *Aquac. Res.* **2008**, *39*, 1450–1460. [\[CrossRef\]](#)
- Daudpota, A.M.; Kalhor, I.B.; Shah, S.A.; Kalhor, H.; Abbas, G. Effect of stocking densities on growth, production and survival rate of red tilapia in hapa at fish hatchery Chilya Thatta, Sindh, Pakistan. *J. Fish.* **2014**, *2*, 180–186. [\[CrossRef\]](#)
- Gao, G.; Xiao, K.; Chen, M. An intelligent IoT-based control and traceability system to forecast and maintain water quality in freshwater fish farms. *Comput. Electron. Agric.* **2019**, *166*, 105013. [\[CrossRef\]](#)
- Ani, J.S.; Manyala, J.O.; Masese, F.O.; Fitzsimmons, K. Effect of stocking density on growth performance of monosex Nile Tilapia (*Oreochromis niloticus*) in the aquaponic system integrated with lettuce (*Lactuca sativa*). *Aquac. Fish.* **2022**, *7*, 328–335. [\[CrossRef\]](#)
- Zambrano, A.F.; Giraldo, L.F.; Quimbayo, J.; Medina, B.; Castillo, E. Machine learning for manually-measured water quality prediction in fish farming. *PLoS ONE* **2021**, *16*, E0256380. [\[CrossRef\]](#)
- Palani, S.; Liong, S.Y.; Tkalich, P. An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **2008**, *56*, 1586–1597. [\[CrossRef\]](#)
- Castrillo, M.; López García, A. Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. *Water Res.* **2020**, *172*, 115490. [\[CrossRef\]](#)
- Anand, M.V.; Sohitha, C.; Saraswathi, G.N.; Lavanya, G.V. Water quality prediction using CNN. *J. Phys. Conf. Ser.* **2023**, *2428*, 012051. [\[CrossRef\]](#)
- Ye, B.; Cao, X.; Liu, H.; Wang, Y.; Tang, B.; Chen, C.; Chen, Q. Water chemical oxygen demand prediction model based on the CNN and ultraviolet-visible spectroscopy. *Front. Environ. Sci.* **2022**, *10*, 1027693. [\[CrossRef\]](#)
- Hu, Z.; Zhang, Y.; Zhao, Y.; Xie, M.; Zhong, J.; Tu, Z.; Liu, J. A water quality prediction method based on the Deep LSTM network considering correlation in smart mariculture. *Sensors* **2019**, *19*, 1420. [\[CrossRef\]](#)
- Liu, P.; Wang, J.; Sangaiah, A.K.; Xie, Y.; Yin, X. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability* **2019**, *11*, 2058. [\[CrossRef\]](#)
- Ahmed, U.; Mumtaz, R.; Anwar, H.; Shah, A.A.; Irfan, R.; García-Nieto, J. Efficient water quality prediction using supervised machine learning. *Water* **2019**, *11*, 2210. [\[CrossRef\]](#)
- Juna, A.; Umer, M.; Sadiq, S.; Karamti, H.; Eshmawi, A.A.; Mohamed, A.; Ashraf, I. Water quality prediction using KNN imputer and multilayer perceptron. *Water* **2022**, *14*, 2592. [\[CrossRef\]](#)
- Li, T.; Lu, J.; Wu, J.; Zhang, Z.; Chen, L. Predicting aquaculture water quality using machine learning approaches. *Water* **2022**, *14*, 2836. [\[CrossRef\]](#)
- Wang, X.; Li, Y.; Qiao, Q.; Tavares, A.; Liang, Y. Water quality prediction based on machine learning and comprehensive weighting methods. *Entropy* **2023**, *25*, 1186. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cojbasic, S.; Dmirasinovic, S.; Kostic, M.; Sekulic, M.T.; Radonic, J.; Dodig, A.; Stojkovic, M. Application of machine learning in river water quality management: A review. *Water Sci. Technol.* **2023**, *88*, 2297–2308. [\[CrossRef\]](#)
- da Silva, L.F.B.A.; Yang, Z.; Pires, N.M.M.; Dong, T.; Teien, H.C.; Storebakken, T.; Salbu, B. Monitoring aquaculture water quality: Design of an early warning sensor with *Aliivibrio fischeri* and predictive models. *Sensors* **2018**, *18*, 2848. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, F.; Du, Y.; Qiu, T.; Xu, Z.; Zhou, L.; Xu, J.; Sun, M.; Li, Y.; Sun, J. Design of an intelligent variable-flow recirculating aquaculture system based on machine learning methods. *Appl. Sci.* **2021**, *11*, 6545. [\[CrossRef\]](#)

22. Yang, J.; Jia, L.; Guo, Z.; Shen, Y.; Li, X.; Mou, Z.; Yu, K.; Lin, J.C.W. Prediction and control of water quality in Recirculating Aquaculture System based on hybrid neural network. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106002. [[CrossRef](#)]
23. Wu, J.; Wang, Z. A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water* **2022**, *14*, 610. [[CrossRef](#)]
24. Zhou, S.; Song, C.; Zhang, J.; Chang, W.; Hou, W.; Yang, L. A hybrid prediction framework for water quality with integrated W-ARIMA-GRU and LightGBM methods. *Water* **2022**, *14*, 1322. [[CrossRef](#)]
25. Chen, H.; Yang, J.; Fu, X.; Zheng, Q.; Song, X.; Fu, Z.; Wang, J.; Liang, Y.; Yin, H.; Liu, Z.; et al. Water quality prediction based on LSTM and attention mechanism: A case study of the Burnett River, Australia. *Sustainability* **2022**, *14*, 13231. [[CrossRef](#)]
26. Cai, H.; Zhang, C.; Xu, J.; Wang, F.; Xiao, L.; Huang, S.; Zhang, Y. Water quality prediction based on the KF-LSTM encoder-decoder network: A case study with missing data collection. *Water* **2023**, *15*, 2542. [[CrossRef](#)]
27. Farzana, S.Z.; Paudyal, D.R.; Chadalavada, S.; Alam, M.J. Prediction of water quality in reservoirs: A comparative assessment of machine learning and deep learning approaches in the case of Toowoomba, Queensland, Australia. *Geosciences* **2023**, *13*, 293. [[CrossRef](#)]
28. APHA. *Standard Methods for the Examination of Water and Wastewater*, 20th ed.; American Public Health Association, American Water Works Association, Water Environment Federation: Washington, DC, USA, 2005.
29. Kolding, J.; Haug, L.; Stefansson, S. Effect of ambient oxygen on growth and reproduction in Nile tilapia (*Oreochromis niloticus*). *Can. J. Fish. Aquat.* **2008**, *65*, 1413–1424. [[CrossRef](#)]
30. Tran-Duy, A.; van Dam, A.A.; Schrama, J.W. Feed intake, growth and metabolism of Nile tilapia (*Oreochromis niloticus*) in relation to dissolved oxygen concentration. *Aquac. Res.* **2012**, *43*, 730–744. [[CrossRef](#)]
31. Azaza, M.S.; Dhraief, M.N.; Kraïem, M. Effect of water temperature on growth and sex ratio of juvenile Nile tilapia *Oreochromis niloticus* (Linnaeus) reared in geothermal waters in southern Tunisia. *J. Therm. Biol.* **2008**, *33*, 98–105. [[CrossRef](#)]
32. Lawson, T.B. *Fundamentals of Aquacultural Engineering*; Chapman & Hall: Orange, CA, USA, 1995.
33. El-Sherif, M.S.; El-Feky, A.M.I. Performance of Nile tilapia (*Oreochromis niloticus*) fingerlings I. Effect of pH. *Int. J. Agric. Biol.* **2009**, *11*, 297–300.
34. Hargreaves, J.A.; Tucker, C.S. *Managing Ammonia in Fish Ponds*; Southern Regional Aquaculture Center: Stoneville, MS, USA, 2004.
35. Stone, N.M.; Thomforde, H.K. *Understanding Your Fish Pond Water Analysis Report*; Cooperative Extension Program, University of Arkansas at Pine Bluff: Pine Bluff, AR, USA, 2004.
36. Boyd, C.E.; Tucker, C.S. *Pond Aquaculture Water Quality Management*; Springer: New York, NY, USA, 2012.
37. Boyd, C.E. *Water Quality Management for Pond Fish Culture*; Elsevier: Amsterdam, The Netherlands, 1982.
38. Wahab, M.A.; Ahmed, Z.F.; Islam, M.A.; Haq, M.S.; Rahmatullah, S.M. Effects of introduction of common carp, *Cyprinus carpio* (L.), on the pond ecology and growth of fish in polyculture. *Aquac. Res.* **1995**, *26*, 619–628. [[CrossRef](#)]
39. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
40. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
41. Lipton, Z.C.; Kale, D.C.; Elkan, C.P.; Wetzell, R.C. Learning to Diagnose with LSTM Recurrent Neural Networks, 2015. Available online: <https://arxiv.org/abs/1511.03677> (accessed on 6 January 2024).
42. Donahue, J.; Hendricks, L.A.; Rohrbach, M.; Venugopalan, S.; Guadarrama, S.; Saenko, K.; Darrell, T. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 677–691. [[CrossRef](#)] [[PubMed](#)]
43. Feizollah, A.; Ainin, S.; Anuar, N.B.; Abdullah, N.A.B.; Hazim, M. Halal products on twitter: Data extraction and sentiment analysis using stack of deep learning algorithms. *IEEE Access* **2019**, *7*, 83354–83362. [[CrossRef](#)]
44. Baek, S.-S.; Pyo, J.; Chun, J.A. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water* **2020**, *12*, 3399. [[CrossRef](#)]
45. Li, P.; Zhang, J.; Krebs, P. Prediction of flow based on a CNN-LSTM combined deep learning approach. *Water* **2022**, *14*, 993. [[CrossRef](#)]
46. Li, Y.; Kong, B.; Yu, W.; Zhu, X. An attention-based CNN-LSTM method for effluent wastewater quality prediction. *Appl. Sci.* **2023**, *13*, 7011. [[CrossRef](#)]
47. Boyd, C.E.; Tucker, C.S. *Handbook for Aquaculture Water Quality*; Craftmaster Printers: Auburn, AL, USA, 2014.
48. Fossmark, R.O.; Vadstein, O.; Rosten, T.W.; Bakke, I.; Košeto, D.; Bugten, A.V.; Helberg, G.A.; Nesje, J.; Jørgensen, N.O.G.; Raspati, G.; et al. Effect of reduced organic matter loading through membrane filtration on the microbial community dynamics in recirculating aquaculture systems (RAS) with Atlantic salmon parr (*Salmo salar*). *Aquaculture* **2020**, *524*, 735268. [[CrossRef](#)]
49. Zhang, X.; Wang, J.; Wang, C.; Li, W.; Ge, Q.; Qin, Z.; Li, J.; Li, J. Effects of long-term high carbonate alkalinity stress on the ovarian development in *Exopalaemon carinicauda*. *Water* **2022**, *14*, 3690. [[CrossRef](#)]
50. Tan, W.K.; Cheah, S.C.; Parthasarathy, S.; Rajesh, R.P.; Pang, C.H.; Manickam, S. Fish pond water treatment using ultrasonic cavitation and advances oxidation processes. *Chemosphere* **2021**, *274*, 129702. [[CrossRef](#)] [[PubMed](#)]
51. Sriyasad, P.; Chitmanat, C.; Whangchai, N.; Promya, J.; Lebel, L. Effect of water de-stratification on dissolved oxygen and ammonia in tilapia pond in Northern Thailand. *Int. Aquat. Res.* **2015**, *7*, 287–299. [[CrossRef](#)]

52. Hardy, L. Modeling nitrogen species as a source of titratable alkalinity and dissolved gas pressure in water. *Appl. Geochem.* **2018**, *98*, 301–309. [[CrossRef](#)]
53. Zhu, S.; Chen, S. The impact of temperature on nitrification rate in fixed film biofilters. *Aquac. Eng.* **2022**, *26*, 221–227. [[CrossRef](#)]
54. Pedersen, O.; Colmer, T.D.; Sand-Jensen, K. Underwater photosynthesis of submerged plants-recent advances and methods. *Front. Plant Sci.* **2013**, *4*, 140. [[CrossRef](#)] [[PubMed](#)]
55. Saalidong, B.M.; Aram, S.A.; Otu, S.; Lartey, P.O. Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems. *PLoS ONE* **2022**, *17*, e0262117. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.