Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:120
https://doi.org/10.28919/cmbn/8841
ISSN: 2052-2541

INVESTIGATING THE CAPABILITY OF LONG SHORT-TERM MEMORY INPUT LAYERS FOR SLIDING WINDOWED DATA FOR ENHANCING WATER QUALITY PARAMETER PREDICTION IN SMALL FISHPONDS

KARLI EKA SETIAWAN[1,*], ERNA FRANSISCA ANGELA SIHOTANG[2], MARVEL MARTAWIDJAJA[3], MUHAMMAD RIZKI NUR MAJIID[4]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[2]Statistics Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[3]Data Science Program, Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[4]Computer Science Department Semarang Campus, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract: This research contribution is a comprehensive discussion of the Long Short-Term Memory (LSTM) capabilities investigation in handling sliding windowed data, especially for enhancing water quality parameter prediction in small fishpond. This research used a public dataset from an aquaponic system containing some information such as water pH, total dissolved solids (TDS), and water temperature. This research experiment on four different input types for our proposed predictive model using LSTM with the best handling way for processing sliding windowed data was input type 4, which used a flattening process with a timestep sequence order, resulting in the

*Corresponding author
E-mail address: karli.setiawan@binus.ac.id
Received August 16, 2024

lowest error of 31.3912 in MAE and 87.3414 for predicting TDS, the lowest error of 0.1769 in MAE and 0.2375 in RMSE for predicting water temperature, and still good enough for predicting water pH with 0.6367 in MAE and 0.8302 in RMSE.

Keywords: deep learning; water quality prediction; aquaponic systems; environmental data prediction; lstm for water quality; sliding window data handling, smart and precision aquaponics.

2020 AMS Subject Classification: 68T01, 68T10, 97P80.

1. INTRODUCTION

Having a fishpond can be a livelihood for income generation and food production, as well as, for some people, a source of leisure, recreation, and even education. The development of some technologies, such as Artificial Intelligence (AI), the Internet of Things (IoT), and cloud computing systems, has also significantly improved the development of fishponds, of which recirculating aquaculture systems (RAS) are one example. RAS is an artificial land-based aquaculture ecosystem where the goals of this system are to maintain appropriate water quality by reusing water and reducing the use of water [1]. In the last forty years, the aquaculture industry has recorded a yearly growth rate of 7% [2]. Meanwhile, from the point of view of the fishery business, the decreasing resource of safe water is one of several big problems due to climate change, contamination, and pollution [3]. Monitoring water is an essential activity in the fishery business to guarantee excellent quality and productivity, as supported by research in the last decade that stated that there is a strong correlation between water quality and fish production [4]. Some researchers reported that the existence of some bacteria, such as Fecal Coliform bacteria, E. coli, the growth of Eutrophication, chemistry element levels such as acidification, pH, chemical pollution, pesticides, herbicides, hardness, and metals, must be monitored from the water quality in fishponds [3]. In conventional ways, water monitoring can be energy-consuming and costly where water samples need to be collected and laboratory-based testing is required [5]. This research performed deep learning (DL) approach to predict the water quality based on the data obtained from IoT device installed in fishpond.

Research was conducted by some researchers, who have a goal of helping farmers manage the quality of water to increase the quality and quantity of shrimp and raise fishery businesses [6]. They utilized a deep learning approach with the long shot-term memory (LSTM) Model and Stacked LSTM to be implemented on their IoT-based monitoring water quality system with a data set containing information such as salinity, temperature, pH, and dissolved oxygen (DO). Their research showed that their single LSTM model outperformed the stacked LSTM and support vector machine (SVM) models. Another similar study using a deep learning approach to monitor water quality in shrimp farming was done by Lin et al. by proposing deep belief networks (DBN) models [7]. Their research was conducted in twelve shrimp ponds in Ningbo, Zhejiang Province, China, using real data from that site containing water heat, pH, salinity level, and DO information. The result showed that the DBN-based Model was efficient and suitable for forecasting water quality. Another IoT-based study was done by some researchers where they utilized autoML algorithm into their aquaponics system to improve plant and fish growth [8]. They monitored some parameters such as temperature, humidity, light intensity and wind speed through think speak cloud platform. They claimed that their proposed model with autoML with gradient boost can give positive impact through the increasing of yield and giving cost-effective for conserving water. Some researchers conducted an investigation using a machine learning (ML) approach to explore the correlation between water quality and harvest performance [9]. The ML models used by them in their exploration were neural networks, SVM, k-nearest neighbors, logistic regression, gaussian naïve bayes, decision tree, random forest, and AdaBoost using a dataset from an Australian prawn farm. In their research, the results showed that DO, salinity, and temperature were the most important and influential parameters for increasing fishpond yield. This research did not determine which ML model was the best, but they concluded that ML approaches have great potential for supporting farmers in developing their fishponds in order to increase their harvest outcomes. There is a study comparison conducted by some engineers in water quality prediction using various deep learning models such as convolutional neural networks (CNN), LSTM, and gated recurrent units (GRU) [10]. The dataset containing information on salinity level, pH, DO, and water heat they used was the data on water quality in aquaculture collected from the Agency for Development of Aquaculture Kerala (ADAK) in Kerala, India. They proposed hybrid deep learning models such as CNN-LSTM and CNN-GRU. Compared to Arima, deep learning baseline models such as LSTM and GRU, and attention-based models such as LSTM and GRU, their proposed CNN-LSTM outperforms all comparison models.

This research is further research from our previous research in enhancing water quality prediction in small fishponds [11]. This research contribution is the investigation of various input for our proposed DL prediction model using LSTM, which resulted a comprehensive analysis of LSTM capability in resulting precise prediction. This research investigated four different inputs for our proposed LSTM model, such as column data format as a common LSTM input, transposed column data format, flattened by parameter type, and flattened by time order which can be seen at Figure 2 in section 2.2.

2. PROPOSED METHODOLOGY

This research focused on observing the capability of the LSTM model in predicting water quality with some parameters such as pH, TDS, and temperature in a small fishpond using four different input formats.

2.1. Data collection and preparation

This study exploited a public dataset from an aquaponic system collected by Siswanto et al. containing pH, TDS, and water temperature data [12]. The dataset was gathered from a one-meter cubic of fish lagoon aquaponic system comprising 30 red tilapia and a hydroponic medium applying the Nutrient Film Technique (NFT) method, according to their research. From January to March 2023, 118,286 rows and five columns of data, including id, created date, pH value, TDS concentration, and water temperature, were recorded by all of the Internet of Things (IoT) devices, which included a NodeMCU ESP8266 microcontroller and three sensors: DS18B20, pH-4502C, and DFROBOT Analog TDS Sensor. The sensors were embedded in the reservoir and wirelessly communicated through the internet. The dataset applied in this research was exactly the same dataset used in our previous research, where we applied some techniques such as duplicate removing and interpolate the missing data in certain data points. Eventually the dataset used in this research is depicted in Figure 1. The characteristics of the dataset are also depicted in Figure 1, with a Pearson Correlation Coefficient (PCC) value that interpreted the correlation between water pH and TDS as having a negligible correlation, between water pH and water temperature as having a negatively moderate correlation, and between TDS and water temperature as having a weak correlation, so that our dataset can be concluded to be a complex dataset [13]. This research split the dataset into two parts, such as the train set and the test set, with ratios of 80% and 20%, respectively. In Figure 1, the train set is illustrated in blue, and the test set is illustrated in orange.

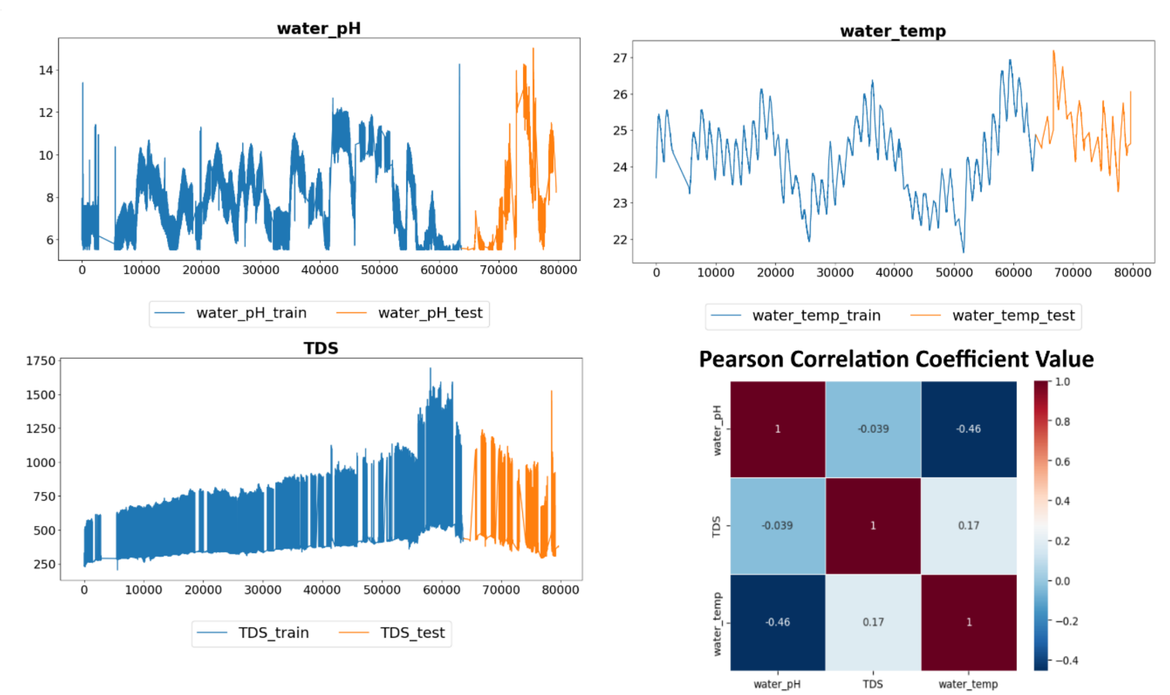

FIGURE 1. Dataset Illustration.

2.2. Investigation scheme and proposed deep learning prediction model

The prediction scenario in this research was set to predict one time in the future based on previous 24 data as depicted in illustration (a) in Figure 2, because the previous 24 data was reasonable enough to make our predictive model to predict the next coming data. Curiosity has guided us by giving us intuition to investigate the various dimensions of the input layer for our LSTM models. The first type of input layer was the common input style for sliding windowed data for LSTM, where each parameter is placed in the same column as in illustration (b) in Figure 2. The second type of our input layer was the transposed version of our first type, where the data in the same timestep were put in the same column as in illustration (c) in Figure 2. The third input layer type was flattened data, where the first sequence was the water pH sequence, the next sequence was TDS, and the last sequence was water temperature as illustration (d) in Figure 2. The fourth of our inputs was the flattened data, where the first sequence was the data in the same timestep, followed by the next timestep, and so on, but for the data in the same timestep, this research decided to put the data sequentially in order: water pH, TDS, and water temperature, as in illustration (e) in Figure 2.

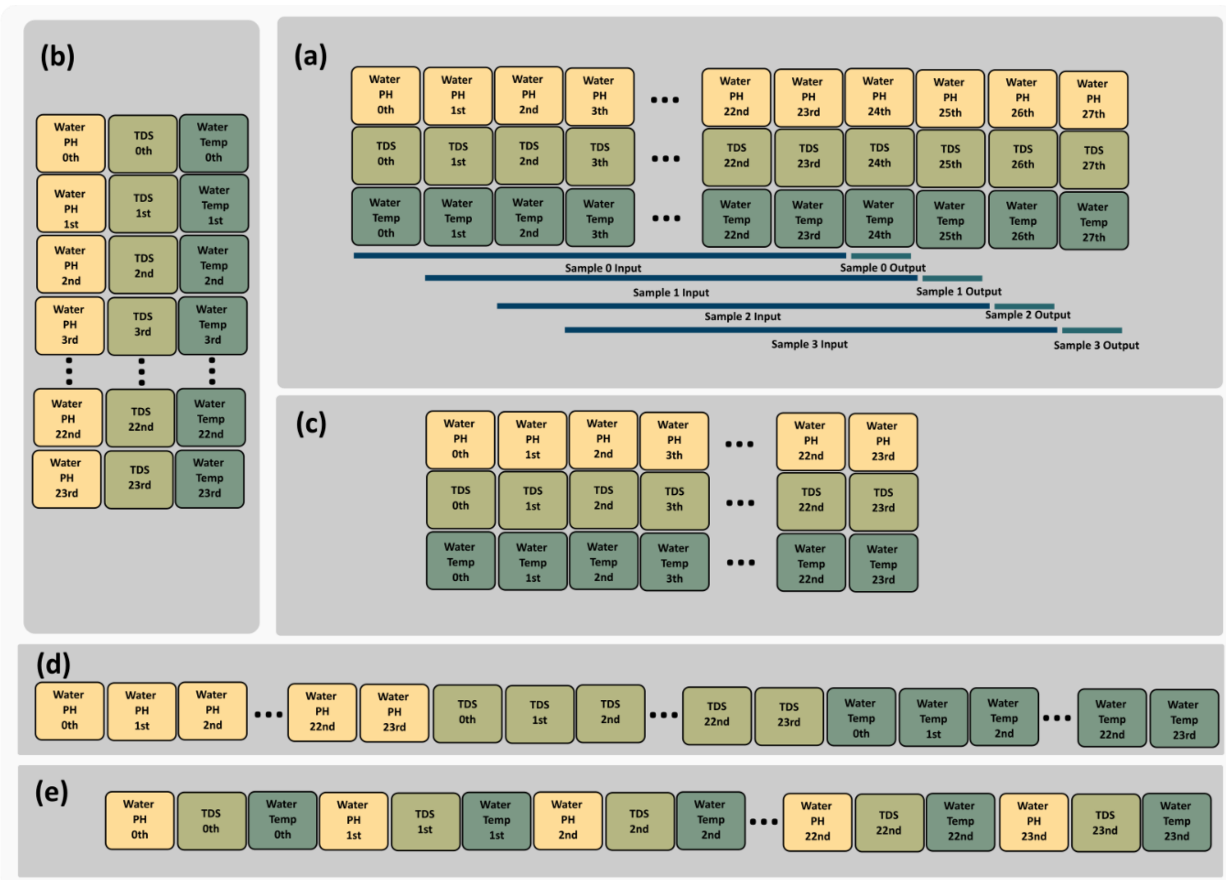

FIGURE 2. Sliding windowed data scenario predicting future data using the previous 24 data points (a), input data for model prediction using the common input style as input type 1 (b), transposed input data scenario as input type 2 (c), flattened input data by parameter type scenario as input type 3 (d), and flattened input data by order time scenario as input type 4 (e).

As in this research has four different input layer types, all predictive deep learning models need to be customized as depicted in Figure 3. Based on Figured 3, the input layer of type 1 and type 2 the number 24 represent the number of timesteps and 3 represent the number of our features such as water pH, TDS, and water temperature, meanwhile on the input layer of type 3 and 4, the number 72 represent the multiplication of 24 timesteps and 3 features. Meanwhile the reason why this research choose LSTM layer was because LSTM have the ability in memorizing temporal information over large number of timesteps that suitable for this research [14], [15]. LSTM is part of Recurrent Neural Network (RNN) family and LSTM works with three gates mechanism such as input gate, output gate, and forget gate.

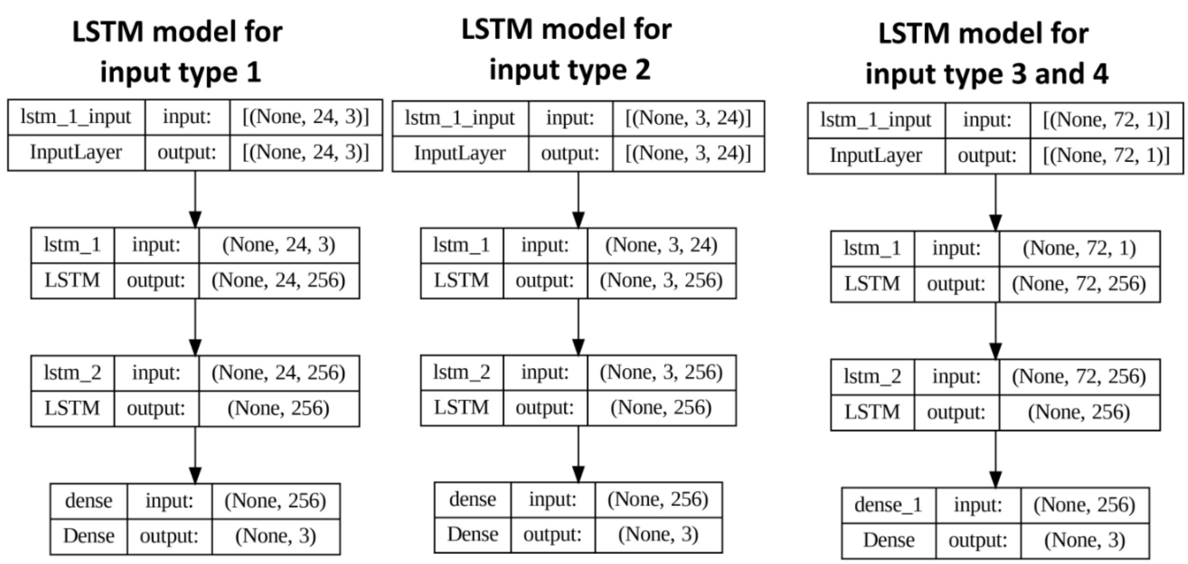

FIGURE 3. Proposed LSTM Prediction Models for each Different Input.

3. MAIN RESULTS

For a fair comparison, this research implemented the Adam optimizer with a 0.00001 learning rate, 256 batch sizes, 256 cell units in the LSTM layer, and 100 training epochs in all the proposed LSTM models. The training process can be seen in Figure 4, where the best process was the LSTM model with input type 4.

TABLE 1. Prediction Testing Result

Parameters/ Features	**Metrics**	**LSTM prediction model result using input type 1**	**LSTM prediction model result using input type 2**	**LSTM prediction model result using input type 3**	**LSTM prediction model result using input type 4**
water pH	MAE	2.1952	**0.3166**	1.1997	0.6367
	RMSE	2.5072	**0.5113**	1.4254	0.8302
TDS	MAE	133.1646	107.8376	71.3682	**31.3917**
	RMSE	206.3211	190.3342	158.4581	**87.3414**
water temperature	MAE	1.0289	0.2553	1.0211	**0.1769**
	RMSE	1.2265	0.3292	1.3933	**0.2375**

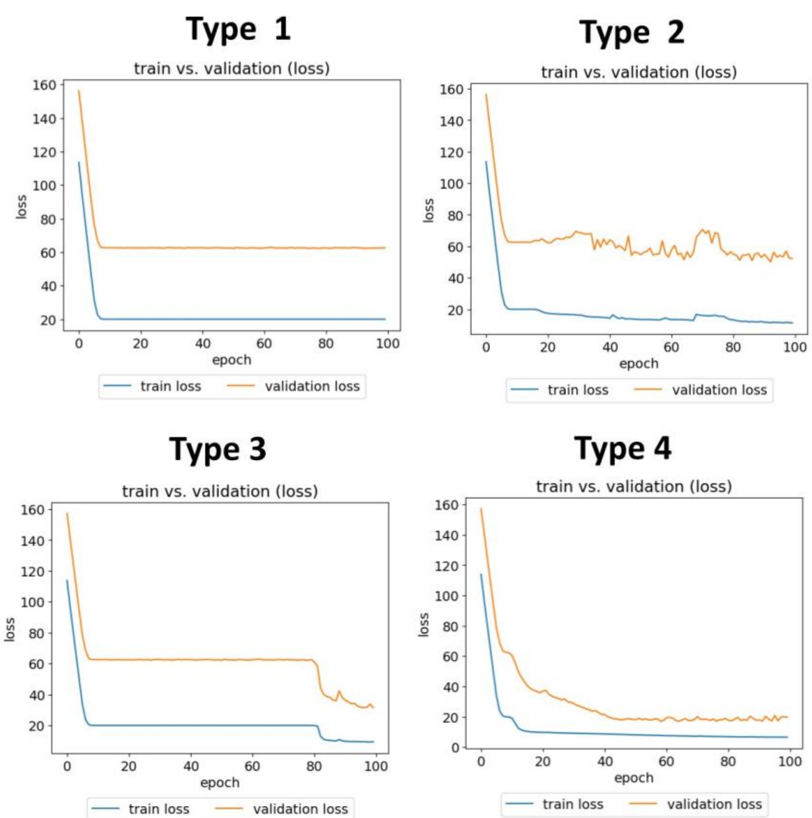

FIGURE 4. Training Process Proposed LSTM Models with Different Input Type.

Overall, based on the testing results shown in Table 1 and Figure 5, it can be concluded that the proposed LSTM model using input type 4 was the best among all other input types. The LSTM model using input type 4 was the best at predicting TDS and water temperature, while the LSTM model using input type 2 was the best at predicting water pH. A quick glance at Figure 5 shows that the LSTM model using input type 1 failed to predict all water quality parameters, including pH, TDS, and temperature, by only resulting in a horizontal straight line. In our analysis, the prediction may not be good enough due to the data tending to be pattern less without containing any pattern such as seasonality, and due to the characteristic of the dataset depicted in PCC values not containing any strong correlation but only containing negligible, weak, and moderate correlation. In our conjecture, the data may not be large enough to make the predictive model learn. We also assumed the dataset used in our research may contain noise because the data was obtained from sensors, which makes predicting water quality more challenging using this dataset.

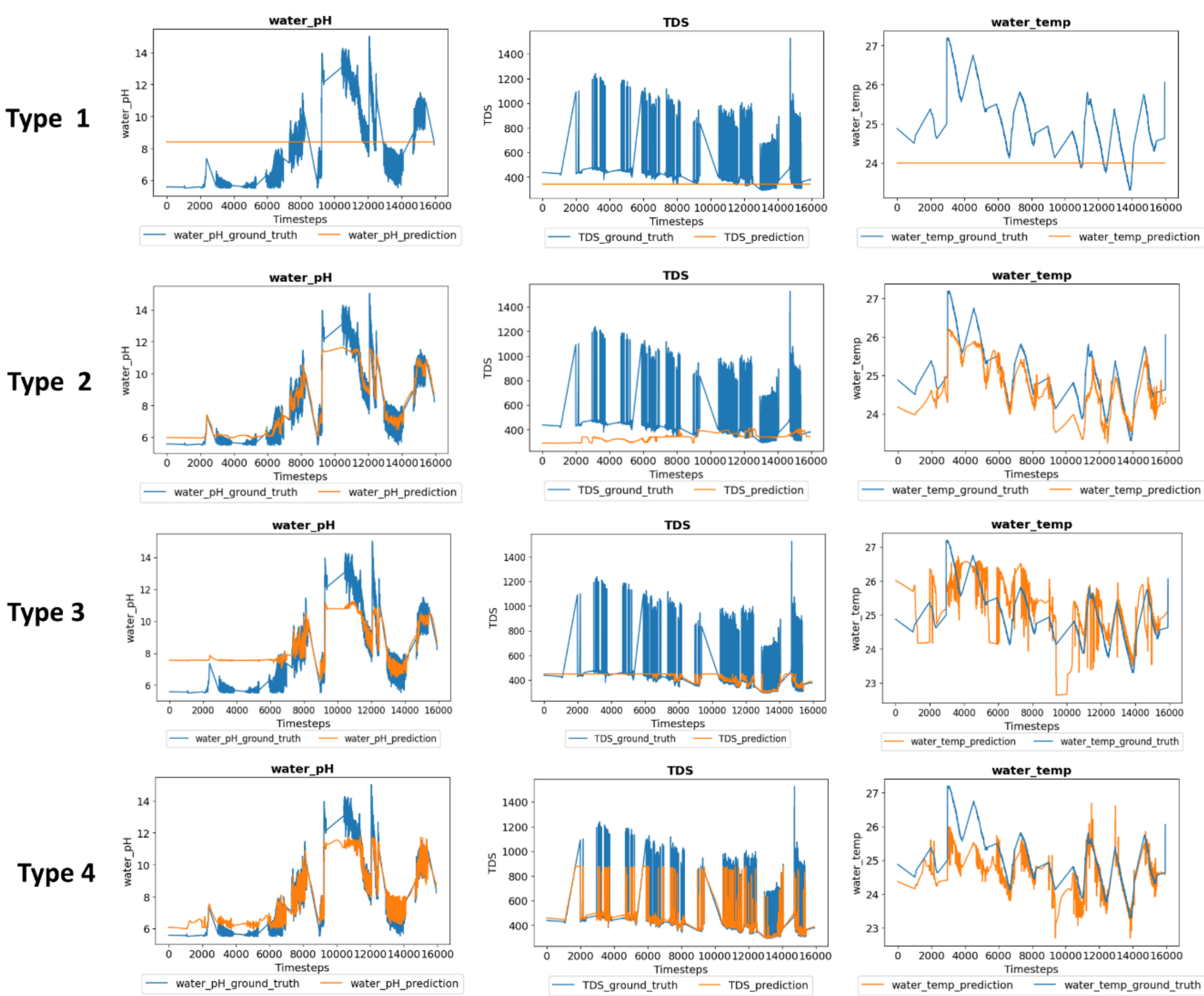

FIGURE 5. Prediction Testing Results.

4. CONCLUSIONS

In conclusion, this research revealed that the LSTM model using input type 4, which means implementing a flattening process with a timestep sequence order, can improve the accuracy of predicting water quality with the lowest error in predicting TDS around 31.3917 in MAE and 87.3414 in RMSE and in predicting water temperature around 0.1769 in MAE and 0.2375 in RMSE, even though predicting water pH around 0.6367 in MAE and 0.8302 in RMSE is not good enough than the LSTM model using input type 2.

There are many ways to improve this research in the future. One of them is for researchers to compare our research using other predictive models from the RNN family, such as GRU, RNN, Bi-LSTM, Bi-GRU, or other deep learning models. We also assumed that the dataset used in this research may contain noise, so the implementation of Kalman filtering may be helpful in handling data with noise and may reduce the error in predicting water quality parameters. In terms of predicting water quality, three variables such as pH, TDS, and temperature may not be enough to depict the quality of water in a small fishpond. A complex dataset containing many other variables may help in resulting a more accurate prediction. Overall, the findings of this work provide a potential basis for developing predictive models in the context of water quality evaluation for tiny fishponds, therefore aiding urban farmers and contributing to the long-term growth of aquaculture techniques in restricted areas.

ACKNOWLEDGEMENT

Bina Nusantara University's internal research grant, particularly from Beginner Research, supports this study.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] P. Lindholm-Lehto, Water quality monitoring in recirculating aquaculture systems, Aquac. Fish Fisher. 3 (2023), 113–131. https://doi.org/10.1002/aff2.102.

[2] X. Su, L. Sutarlie, X.J. Loh, Sensors, biosensors, and analytical technologies for aquaculture water quality, Research 2020 (2020), 8272705. https://doi.org/10.34133/2020/8272705.

[3] M. Manoj, V.D. Kumar, M. Arif, et al. State of the art techniques for water quality monitoring systems for fish ponds using IoT and underwater sensors: a review, Sensors 22 (2022), 2088. https://doi.org/10.3390/s22062088.

[4] S.Y. Zhang, G. Li, H.B. Wu, et al. An integrated recirculating aquaculture system (ras) for land-based fish farming: the effects on water quality and fish production, Aquac. Eng. 45 (2011), 93–102. https://doi.org/10.1016/j.aquaeng.2011.08.001.

[5] C.Z. Zulkifli, S. Garfan, M. Talal, et al. IoT-based water monitoring systems: a systematic review, Water 14 (2022), 3621. https://doi.org/10.3390/w14223621.

[6] N. Thai-Nghe, N. Thanh-Hai, N.C. Ngon, Deep learning approach for forecasting water quality in iot systems. Int J. Adv. Comput. Sci. Appl. 11 (2020), 686–693. https://archimer.ifremer.fr/doc/00646/75836/

[7] Q. Lin, W. Yang, C. Zheng, et al. Deep-learning based approach for forecast of water quality in intensive shrimp culture ponds, Indian J. Fish. 65 (2018), 75-80. https://doi.org/10.21077/ijf.2018.65.4.72559-09.

[8] C.S. Arvind, R. Jyothi, K. Kaushal, et al. Edge computing based smart aquaponics monitoring system using deep learning in IoT environment, in: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, Canberra, ACT, Australia, 2020: pp. 1485–1491. https://doi.org/10.1109/SSCI47803.2020.9308395.

[9] M. Rana, A. Rahman, J. Dabrowski, et al. Machine learning approach to investigate the influence of water quality on aquatic livestock in freshwater ponds, Biosyst. Eng. 208 (2021), 164–175. https://doi.org/10.1016/j.biosystemseng.2021.05.017.

[10] K.P. Rasheed Abdul Haq, V.P. Harigovindan, Water quality prediction for smart aquaculture using hybrid deep learning models, IEEE Access 10 (2022), 60078–60098. https://doi.org/10.1109/access.2022.3180482.

[11] M.R.N. Majiid, K.E. Setiawan, Ensemble shuffling-cnn, attention-alternative model of water quality parameters prediction in small fishponds for urban aquaculture, ICIC Express Lett. 18 (2024), 669-676. https://doi.org/10.24507/icicel.18.07.669.

[12] B. Siswanto, Y. Dani, D. Morika, B. Mardiyana, A simple dataset of water quality on aquaponic fish ponds based on an internet of things measurement device, Data Brief 48 (2023) 109248. https://doi.org/10.1016/j.dib.2023.109248.

[13] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesth. Analg. 126 (2018), 1763–1768. https://doi.org/10.1213/ane.0000000000002864.

[14] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Comparison of deep learning sequence-to-sequence models in predicting indoor temperature and humidity in solar dryer dome, Commun. Math. Biol. Neurosci., 2022 (2022), 98. https://doi.org/10.28919/cmbn/7655.

[15] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Sequence to sequence deep learning architecture for forecasting temperature and humidity inside closed space, in: 2022 10th International Conference on Cyber and IT Service Management, IEEE, Yogyakarta, Indonesia, 2022: pp. 1–7. https://doi.org/10.1109/CITSM56380.2022.9936008.