

Received June 24, 2020, accepted August 5, 2020, date of publication August 17, 2020, date of current version September 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017089

Long Short-Term Memory Recurrent Neural Network for Tidal Level Forecasting

CHENG-HONG YANG^{1,2}, (Senior Member, IEEE), CHIH-HSIEN WU¹, AND CHIH-MIN HSIEH³

¹Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City 80778, Taiwan

²Department of Biomedical Engineering, Kaohsiung Medical University, Kaohsiung City 80708, Taiwan

³Department of Maritime Information and Technology, National Kaohsiung University of Science and Technology, Kaohsiung City 80778, Taiwan

Corresponding author: Chih-Min Hsieh (chihmin_623@yahoo.com.tw)

This work was supported by the Ministry of Science and Technology, China, under Grant 108-2221-E-992-031-MY3.

ABSTRACT Tide is a phenomenon of water level change caused by gravity. Tidal level forecasting is not only a key theoretical topic but also crucial in coastal and ocean engineering applications. The waiting time before a cargo ship enters a port affects the efficiency of cargo transportation, the tidal difference affects the establishment of turbine generators, and an excessive tidal water level reduces vessel safety. With the proliferation of information technology, the application of deep learning models in the analysis and study of hydrological problems has become increasingly common. This study proposed a deep learning model to predict the tidal water level. A forecasting model was developed on the basis of the long short-term memory (LSTM) recurrent neural network for predicting the water levels of 17 harbors in Taiwan. Tidal water level data for 21 years were collected from different observation stations. To objectively evaluate model performance, the developed model was compared with six other forecasting models in terms of the mean absolute percentage error (MAPE) and root mean square error (RMSE) of the forecasting results. The results indicated that the LSTM model had the lowest forecasting error for the tidal water level for up to 30 days. The average MAPE and RMSE values for the developed model were 6.97% and 0.049 m, respectively; thus, the model could effectively reduce the overlapping problems caused by machine learning methods in continuous forecasting.

INDEX TERMS Deep learning, long short-term memory, tidal level forecasting, time series.

I. INTRODUCTION

Tide is mainly affected by celestial gravity, climate, and air pressure. Therefore, tide is a relatively regular phenomenon in which the water level rises and falls. Tidal changes are closely related to human activities, such as marine economic activities, port development, research plans for coastal and port construction projects, and budget control, which are crucial for economic development [1]. Therefore, tide forecasting requires high-precision tools and methods. For shallow watercourses, tidal changes limit the time that large ships can enter and exit a port. When ebb tide occurs, ships cannot enter certain ports with large changes in tidal levels. According to statistics from Taiwan International Ports Corporation, Ltd. [2], from January to November 2019, 71 841 ships entered or exited Taiwanese ports, with a total tonnage of 1 492 676 527 tons. Moreover, cargo throughput reached 174

437 692 metric tons, with a trade value of approximately NT\$286 778 285. Cargo transported by ships significantly influences Taiwan's economy. Tidal changes should be considered when a ship is at a dock. The tightness of the mooring should be considered to avoid collisions or grounding due to cable breakage on account of excessive tide changes. Accurate tide forecasts are critical to the safety of ships entering and leaving ports as well as the efficiency of port transportation.

Although tidal water levels change periodically, nonlinear fluctuations in water level occur according to changes in the terrain, pressure, time, and moon position at a port. Therefore, accurately predicting water levels is difficult. Tidal level data are related to not only the current time but also previous data. Harmonic analysis proposed by Godin in 1972 is a crucial method for predicting tidal water level [3]. Harmonic analysis has been improved over time [4]. The response [5] and continuous wavelet transform methods [6] have also been applied in tidal water level prediction. After 2005, due to

The associate editor coordinating the review of this manuscript and approving it for publication was Qiang Lai.

the rise of deep learning, artificial neural networks (ANNs) have also been applied in tidal water level prediction [7], [8]. Because ANNs have excellent nonlinear problem processing capabilities, they have gradually replaced other methods of tidal water prediction.

Neural networks, also known as perceptual neuron models, were proposed in the 1950s and mimic the human brain tissue and manipulation method [9]. Prior to the 1980s, the expert system was the most popular artificial intelligence method, and neural network theory was still immature. In the 1980s, the expert system encountered bottlenecks. Neural networks have received increasing attention since their proposition by Hopfield *et al.* [10]. New structures and theories of neural networks are still being developed, and their computation speeds are constantly increasing. Thus, the accuracy and use of these networks is increasing. Deep neural networks (DNNs) have been widely used in various fields over the past 10 years [11]. DNNs have been used in fields such as speech recognition [12], image recognition and classification [13], overall survival prediction [14], and time series prediction [15]. In response to the time series problem, DNNs were improved to create recurrent neural networks (RNNs) and long short-term memory (LSTM) networks [16], [17]. In addition to reducing the prediction error, the aforementioned networks solved the overfitting-related problems of DNNs.

In tide analysis, a tide can be regarded as the result of water level change caused by the superposition of the astronomical tide affected by gravity and the nonlinear water level affected by environmental factors. Harmonic analysis was previously the main method used to predict tidal level [18]; however, its disadvantage is its requirement of long-term data to achieve satisfactory accuracy. Long-term tidal prediction is influenced by phenomena such as noises, seasonal effects, missing data, and typhoon-induced surge [19], [20]. Although many neural network structures have been used to solve this problem, most face the same problem [21]–[23]. In other words, the powerful learning ability of neural networks causes the disadvantage of easy overfitting during training. Consequently, the trained models are only suitable for current data. Moreover, they require numerous hyperparameter adjustments, which considerably increase the time required for training.

This study proposes the addition of a network structure based on LSTM to the fully connected layer for solving the aforementioned problems. An LSTM recurrent neural network comprises one control unit (gate) each for forgetting, updating, and output [24]; thus, the LSTM recurrent neural network has a different structure than does the general neural network. Information can be stored long term in the LSTM recurrent neural network. Moreover, meaningless information is forgotten in the LSTM recurrent neural network, which prevents the phenomenon of vanishing or exploding gradient caused by gradient descent. The LSTM recurrent neural network solves the problem of error superposition caused by the application of a neural network to long-term tide prediction.

Error superposition leads to a large misalignment in predicted result. According to the extended application of DNNs and satisfactory results obtained in time series prediction, the tidal water level prediction performance of the LSTM was compared with that of five other methods commonly used in time series prediction.

II. METHOD

A. NORMALIZATION

Because the conversion function of the LSTM recurrent neural network is a hyperbolic tangent function (tanh function: $e^x - e^{-x}/e^x + e^{-x}$, whose function value ranges from 1 to -1), the training and testing set data are normalized. The normalization equation is as follows:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where $\max(x)$ and $\min(x)$ are the maximum and minimum values in the training and testing sets, respectively, x is the input value, and z is the converted value of x .

B. LSTM

LSTM networks are a variant of RNNs that have been applied in fields such as biomedical science [25], speech recognition [26], sentiment analysis [27], and image classification [28]. However, LSTM recurrent neural networks have not yet been applied in tidal water level forecasting.

The LSTM recurrent neural network provides a solution to the time series problem. Although the backward transfer neural network can establish the correlation between data points through feedback, the function of LSTM is more complete because of the network structure of LSTM units and the forget gates added to the system. These gates can record past crucial features during training and select unimportant features to forget according to weights.

One-step-ahead prediction of the tidal water level time series requires not only current tidal data but also previous data. The RNN model has a self-feedback mechanism in the hidden layer. Because of this mechanism, the RNN model has an advantage in managing long-term dependence problems; however, difficulties remain in practical application [29]. To solve the RNN's problem of gradient vanishing, Hochreiter and Schmidhuber proposed the LSTM model in 1997 [24]. This model was recently improved by Graves [30]. An LSTM unit consists of a memory cell that stores information and is updated by three specialized gates: the input, forget, and output gates. The structure of an LSTM unit is displayed in Fig. 1.

At time t , x_t is the input data of the LSTM cell, h_{t-1} is the output of the LSTM cell at time $t-1$, c_t is the value of the memory cell, and h_t is the output of the LSTM cell. The calculation process of the LSTM unit comprises the following steps:

- (1) First, the value of the candidate memory cell \tilde{c}_t is calculated using (1). In (1), W_c is the weight matrix and

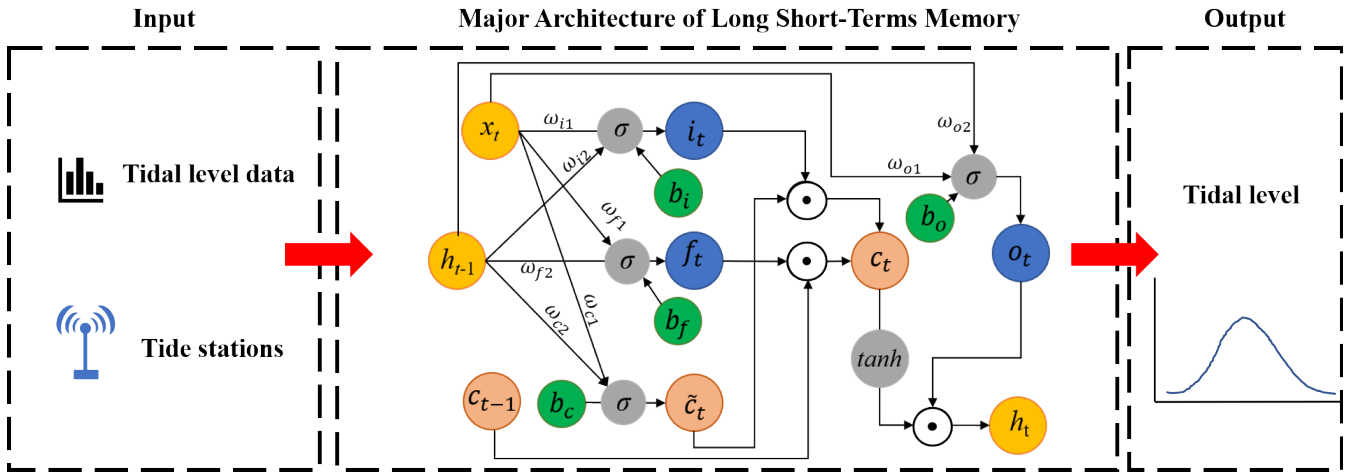


FIGURE 1. Long short-term memory recurrent neural network architecture for data forecasting. Above \odot is Hadamard product. σ represents the sigmoid function. \tanh represents the hyperbolic function.

b_c is the bias.

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (1)$$

- (2) The value of the input gate i_t is then calculated using (2). The input gate controls the updating of the current input data to the state value of the memory cell. In (2), σ is the sigmoid function, W_i is the weight matrix, and b_i is the bias.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

- (3) The value of the forget gate f_t is calculated using (3). The forget gate controls the updating of the historical data to the state value of the memory cell. In (3), W_f is the weight matrix and b_f is the bias.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

- (4) The value of the current moment memory cell c_t is calculated using (4), in which c_{t-1} is the state value of the last LSTM unit.

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (4)$$

- (5) The value of the output gate o_t is calculated using (5). The output gate controls the output of the state value of the memory cell. In (5), W_o is the weight matrix and b_o is the bias.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

- (6) Finally, the output of the LSTM unit h_t is calculated using (6).

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

The three control gates and memory cell of the LSTM unit allow it to maintain, read, reset, and update long-time information easily. Because of the sharing mechanism of the LSTM internal parameters, the dimensions of the output can be controlled by setting the dimensions of the weight matrix. In the LSTM unit, a long delay occurs between input

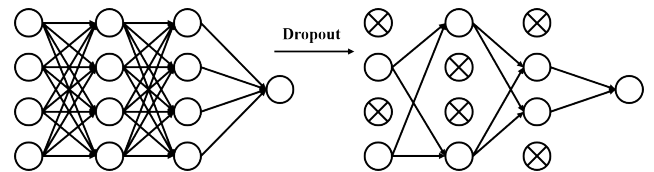


FIGURE 2. Flowchart of dropout.

and feedback. The gradient neither explodes nor disappears because the internal state of the memory cell in the LSTM architecture maintains a constant error flow.

C. DROPOUT

During neural network training, preventing overfitting is crucial. Hinton proposed the dropout method to prevent overfitting in neural network training [31]. During the training process between layers in a neural network, some neurons are randomly dropped from the network with a certain probability, as displayed in Fig. 2.

D. PROPOSED METHOD

To establish a tidal water level forecasting model, this study used two layers of LSTM units. Because the output of the LSTM unit is a multidimensional vector, three fully connected layers are used in each unit. The predicted value is output at the sixth layer. The complete network model structure is displayed in Fig. 3.

Neural network hyperparameters lead how the network functions, and further determine its accuracy and validity. In order to achieve successful performance in each problem domain, LSTM hyperparameters must be adjusted. The hyperparameters of LSTM include number of hidden layers, number of neurons, learning rate, activation function, batch size, epoch and loss function. In this study, depending on how the tuned hyperparameter values affect model performance, LSTM hyperparameters optimization were adjusted manually by the experts. The following hyperparameters settings were obtained, the learning rate is 0.0001, and there are

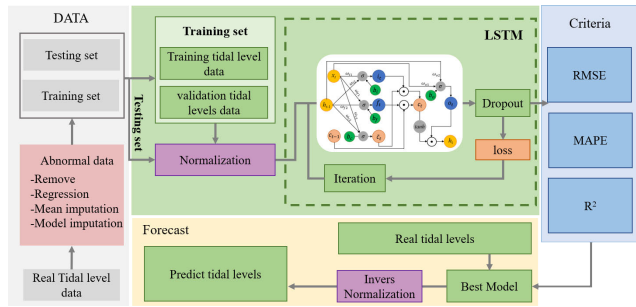


FIGURE 3. Flowchart of the LSTM recurrent neural network.

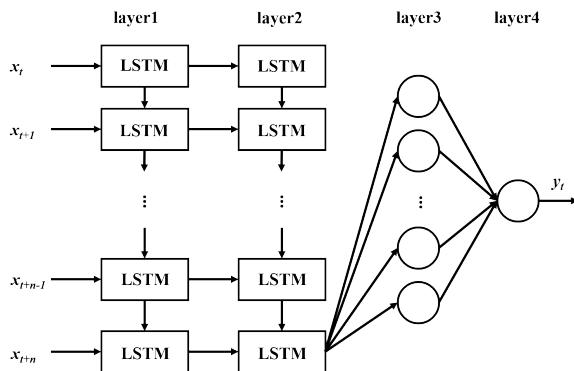


FIGURE 4. Network structure of the proposed model.

three hidden layers, the number of neurons are 200, 150, 50, respectively.

To extract features (Fig. 4), layer 3 has 50 neurons and the output layer has only one neuron, which is also used as the output of the predicted value y_t . All the activation functions use rectified linear units (ReLU) because in addition to the simple calculation process, ReLUs can perform gradient descent and reverse transfer efficiently, thereby preventing gradient explosion and vanishing.

E. STATISTICAL METHOD

1) AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL

The autoregressive integrated moving average (ARIMA) model was proposed by Box and Jenkins in 1976 [32] and is also known as the Box–Jenkins model. In the ARIMA method, several fragments formed after a time series has passed are used as input. Moreover, regression analysis is performed to establish a mathematical forecasting model, which is often used for the prediction of short-term economic trends.

2) TRIGONOMETRIC SEASONALITY, BOX–COX TRANSFORMATION, ARMA ERRORS, AND TREND SEASONAL COMPONENTS MODEL

The trigonometric seasonality, Box–Cox transformation, ARMA errors, and trend seasonal components (TBATS) model, which was proposed by Livera in 2011 [33], is a new method that combines trigonometric seasonality, Box–Cox transformation, ARMA errors, trend, and

seasonal components. The TBATS model is based on the Exponential smoothing. It can predict whether seasonal data exists and can analyze this data. Although a combination of multiple models can provide highly accurate results, considerable training time is required for such a combination, which results in slow calculations.

F. MACHINE LEARNING

1) SUPPORT VECTOR REGRESSION

Support vector regression (SVR) was proposed by Vapnik *et al.* in 1997 [34]. The SVR algorithm includes functions such as the insensitive loss and penalty factor functions; thus, it is more robust than is the support vector machine algorithm [35], [36]. After the SVR algorithm projects the data onto a high-dimensional hyperplane, the total distance from each point to the hyperplane is calculated. If a hyperplane is identified using the minimum total distance, this hyperplane is the solution. Prior to the emergence of deep learning, SVR was the most common method for predicting entire time series.

2) PARTICLE SWARM OPTIMIZATION-BASED SUPPORT VECTOR REGRESSION

Particle swarm optimization (PSO), proposed by Kennedy and Eberhart in 1995, originated from analysis of bird flight when foraging. In particular, birds provide constant updates on the location of insect food sources to the entire group, and the group decides on the optimal feeding ground, akin to solving an optimization problem [37]. In SVR modeling, the parameter settings affect the performance of a forecast time series, as mentioned in the preceding discussion on PSO. The crucial parameters are the regularization parameter (C), bandwidth of the kernel function (σ), and tube size of the ε -insensitive loss function (ε). Inappropriately selected parameter values can result in either overfitting or underfitting [38]. Consequently, selecting the optimal parameters is crucial when employing SVR to forecast a time series. Liu *et al.* used PSO-based SVR to forecast tourist arrivals [38].

3) ARTIFICIAL NEURAL NETWORK

The backpropagation network (BPN) proposed by Hinton is the most commonly used supervised-learning ANN model [39]. The BPN is an optimization algorithm that combines the backward pass, gradient descent [40], and chain rules in calculus. The gradient descent method is used to advance from the initial position of the parameter to the steepest downhill direction and to update the parameter position. The slope information is obtained using the derivative function (to calculate the function slope). Gradient descent uses this characteristic to optimize the cost function.

4) CONVOLUTION NEURAL NETWORK

Convolution neural network (CNN) was proposed by LeCun in 1989 [41]. CNN is a multilayer neural network structure

simulating the operation mechanism of a biological vision system. It is a neural network composed of a multilayer convolution layer and a descending sampling layer. Moreover, CNN can obtain useful feature descriptions from original data, which is an effective method to extract features from data [42].

CNN is a special type of NN that introduces convolution and pooling operations to generate deep features, thus improving the network's ability to recognize patterns. Many studies have shown that CNNs are effective tools for handling complex tasks, such as image recognition [43], text recognition [44], and video recognition [45]. Thus, because of the properties of its convolutional layers, CNN is plausibly appropriate for a seasonal time series with trends. A CNN framework containing convolutional layers, pooling layers, and fully connected layers for time series classification was designed in [46]. This framework performs well in discovering and extracting the internal structure of data.

G. PERFORMANCE CRITERIA

The root mean square error (RMSE) and mean absolute percentage error (MAPE), which are common statistical variables, are used to compare the deviation of the real value from the forecasted value for evaluating the forecasting performance of DNNs. The RMSE, the MAPE, and R^2 are expressed in (7), (8) and (9), respectively.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2} \quad (7)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - f_i}{y_i} \right| \times 100\% \quad (8)$$

$$R^2 = \left[\frac{\sum_{i=1}^n (\eta_{io} - \bar{\eta}_o)(\eta_{im} - \bar{\eta}_M)}{\sqrt{\sum_{i=1}^n (\eta_{io} - \bar{\eta}_o)^2 \sum_{i=1}^n (\eta_{im} - \bar{\eta}_M)^2}} \right]^2 \quad (9)$$

where η_{io} is the observed tidal water level at the i th time step, η_{im} is the corresponding simulated tidal water level, n is the number of time steps, $\bar{\eta}_o$ is the mean of observational values, and $\bar{\eta}_M$ is the mean value of the simulations.

Table 1 indicates that most ports had a low skewness in their tidal data and close to Gaussian distribution. However, the absolute coefficients of variation (COV) of ports 1116, 1246, 1366, 1396, and 1436 were considerably larger than those of other ports. The higher the absolute COV, the greater the degree of data dispersibility is. The average difference between the highest and lowest observed tidal water level reached 7.36 m. Such extreme data can be interpreted as water level rise caused by changes in the terrain natural disasters such as typhoons and tsunamis, and sudden storms.

III. RESULTS

A. DATASETS

In this study, the data of 17 tide stations in Taiwan were analyzed. The names of the 17 stations are Keelung (1516),

TABLE 1. Tidal water level statistics for 17 ports in Taiwan from 1998 to 2018.

| Port | SD | Min | Max | COV(%) |
|------|------|-------|------|--------|
| 112 | 1.24 | -2.70 | 3.75 | 11.05 |
| 113 | 0.34 | -1.49 | 1.47 | -2.31 |
| 1102 | 0.74 | -2.89 | 2.65 | 2.70 |
| 1116 | 0.94 | -2.54 | 2.77 | 27.77 |
| 1166 | 0.57 | -2.05 | 2.00 | 2.03 |
| 1176 | 0.45 | -1.12 | 1.47 | 1.85 |
| 1246 | 0.41 | -1.46 | 1.39 | 33.66 |
| 1256 | 0.41 | -1.70 | 1.48 | -8.90 |
| 1276 | 0.47 | -1.37 | 1.80 | 3.49 |
| 1356 | 0.73 | 0.05 | 2.03 | -8.01 |
| 1366 | 0.58 | -1.92 | 1.88 | 39.73 |
| 1396 | 0.41 | -5.80 | 1.60 | -27.10 |
| 1436 | 1.38 | -3.87 | 3.49 | -31.29 |
| 1486 | 0.26 | -0.69 | 1.75 | 1.66 |
| 1496 | 0.29 | -0.82 | 2.28 | 1.42 |
| 1516 | 0.29 | -1.25 | 1.31 | -11.84 |
| 1586 | 0.40 | -1.20 | 1.57 | 4.77 |

SD, standard deviation; COV, coefficient of variation.

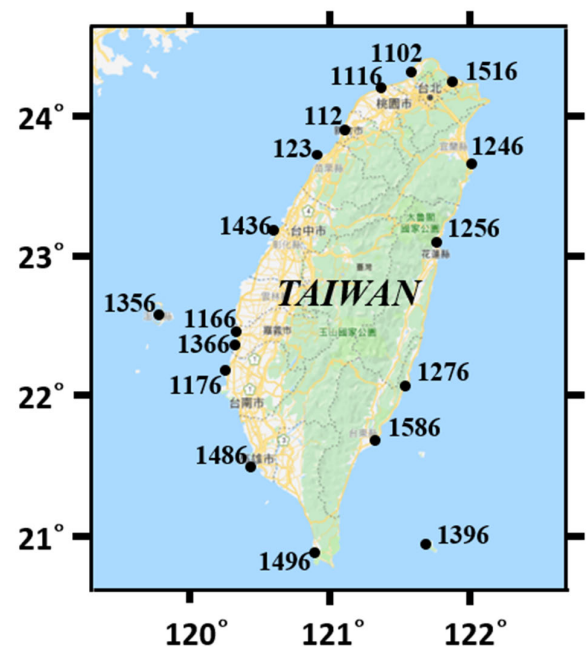


FIGURE 5. Map of tide stations in Taiwan. Keelung (1516), Tamsui (1102), Zhuwei (1116), Hsinchu (112), Waipu (113), Taichung (1436), Wengang (1366), Dongshi (1166), Jiangjun (1176), Kaohsiung (1486), Xunguangzui (1496), Su-ao (1246), Hualien (1256), Chenggong (1276), Fugang (1586), Lanyu (1396), Penghu (1356).

Tamsui (1102), Zhuwei (1116), Hsinchu (112), Waipu (113), Taichung Port (1436), Wengang (1366), Dongshi (1166), Jiangjun (1176), Kaohsiung (1486), Xunguangzui (1496), Su-ao (1246), Hualien (1256), Chenggong (1276), Fugang (1586), Lanyu (1396), and Penghu (1356). Tidal water level data for 21 years were obtained after removing the influence of air pressure. In this study, the LSTM model was used to predict the tidal water level. Study data for 1998–2018 were obtained from the Central Weather Bureau of Taiwan (Fig. 5).

The tide stations were divided into two categories: stations using sonic tide gauges and those using pressure tide gauges for tidal level detection. Sonic tide gauges use microprocessor-based technologies to collect sea level data, whereas pressure tide gauges use mechanical floats and recorders. Modern monitoring stations use advanced acoustics and electronics. Currently used recorders send an audio signal down a half-inch-wide “sounding tube” and measure the time the reflected signal requires to travel back from the water’s surface. Pressure tide gauges are installed on the sea bed. The instantaneous wave height is calculated according to the water pressure change caused by wave height. However, a pressure tide gauge is suitable only for shallow sea areas (approximately 20 m deep). The change in water pressure caused by lifting is too small to obtain accurate data.

B. DATA PREPROCESSING

In this study, 21 years of data were divided into training and testing sets. The training set comprised data from January 1, 1998 to November 30, 2018, and the testing set comprised data from December 1 to December 30, 2018. To reduce the sample size without considerably affecting prediction accuracy, the training and testing sets were normalized such that all data points were within the (0, 1) interval. In addition to achieving standardization, normalization can increase the speed of neural network training. When normalization is performed, convergence can be achieved relatively rapidly, and the possibility of falling into the local optimal solution is relatively low [47]. Moreover, to eliminate the effect of missing values on the prediction, missing values were deleted to prevent the learning of incorrect features during the learning process.

C. ANALYSIS OF FORECASTING RESULTS

The evaluation of the seven methods indicated that the LSTM and ANN models had similar accuracy and superior performance compared with the conventional statistical models. Fig. 6 displays the observed tidal values and the forecasted tidal level values generated by the optimal LSTM, CNN, ANN, PSOSVR, SVR, TBATS, and ARIMA models during prediction. The figure indicates that the LSTM and ANN models had similar accuracies. Compared with the other models, the statistical models had more decentralized predictions. The prediction results indicated that an increase in the prediction interval reduced model accuracy (Table 2). With an increase time in the prediction interval, the R^2 value of the LSTM model reduced from 0.993 to 0.987, that of the CNN model reduced from 0.982 to 0.953, that of the ANN model reduced from 0.984 to 0.946, that of the PSOSVR model reduced from 0.997 to 0.970, that of the SVR model reduced from 0.889 to 0.870, that of the TBATS model reduced from 0.975 to 0.938, and that of the ARIMA model reduced from 0.814 to 0.804. The LSTM model outperformed the ARIMA, TBATS, SVR, PSOSVR, ANN, and CNN models at all five prediction intervals. The LSTM model exhibited the best prediction results, and its scattered point distribution was the closest to the regression line.

TABLE 2. R^2 values of the ARIMA, TBATS, SVR, PSOSVR, ANN, CNN, and LSTM models.

| Prediction interval | 144h | 288h | 432h | 576h | 720h |
|---------------------|-------|-------|-------|-------|-------|
| ARIMA | 0.814 | 0.810 | 0.809 | 0.807 | 0.808 |
| TBATS | 0.889 | 0.886 | 0.877 | 0.870 | 0.870 |
| SVR | 0.889 | 0.886 | 0.877 | 0.870 | 0.773 |
| PSOSVR | 0.997 | 0.996 | 0.989 | 0.980 | 0.970 |
| ANN | 0.984 | 0.983 | 0.953 | 0.951 | 0.951 |
| CNN | 0.982 | 0.975 | 0.963 | 0.954 | 0.953 |
| LSTM | 0.993 | 0.990 | 0.989 | 0.988 | 0.987 |

ARIMA, autoregressive integrated moving average; SVR, support vector regression; TBATS, Box-Cox transformation, ARMA errors, and trend seasonal components; PSOSVR, particle swarm optimization-based support vector regression; ANN, artificial neural network; CNN, convolution neural network; LSTM, long short-term memory.

The parameter R^2 represents the interpretability of the forecasting model for the forecasting result. The higher the R^2 value, the higher is the interpretability and the higher is the model accuracy. As the forecast time increases, the accuracy of the time series forecast decreases; however, exceptions exist to this trend. Because of the increase in the amount of data, a small amount of accurate data occasionally causes R^2 to increase (e.g., the performance of the ARIMA models at 576 and 720 h; Table 2).

This study used LSTM recurrent neural networks for the long-term prediction of tidal water levels. To demonstrate the forecasting performance of our LSTM recurrent neural networks, we evaluated LSTM’s forecasting performance against those of two statistical methods (ARIMA and TBATS methods) and four machine learning methods (SVR, PSOSVR, ANN, and CNN). For all methods, the MAPE and RMSE were used as indicators of forecasting performance; their averages are presented in Table 3. The LSTM method’s MAPE value was 86% and 62% lower than those of ARIMA and TBATS, respectively; the LSTM method’s RMSE value was 90% and 79% lower than those of ARIMA and TBATS, respectively. The LSTM predicted tidal water level significantly better than ARIMA and TBATS did (Fig. 7). As for the machine learning methods, the LSTM method’s MAPE value was 84%, 42%, 48%, and 40% lower than those of the SVR, PSOSVR, ANN, and CNN approaches, respectively; the LSTM method’s RMSE value was 75%, 55%, 65%, and 51% lower than those of SVR, PSOSVR, ANN, and CNN, respectively. These results indicated that the LSTM model significantly outperformed the machine learning methods, which are commonly used in time series predictions of tidal levels.

Fig. 8 illustrates the changes in the MAPE and RMSE of the seven methods. The red solid line represents the RMSE, and the blue solid line represents the MAPE. The error value increased with the number of forecast days. Figs. 7(b), (c), (d), and (f) indicate a significant increase in error value when the total forecasting length was between 40% and 80% (approximately 12–24 days). The growth rate

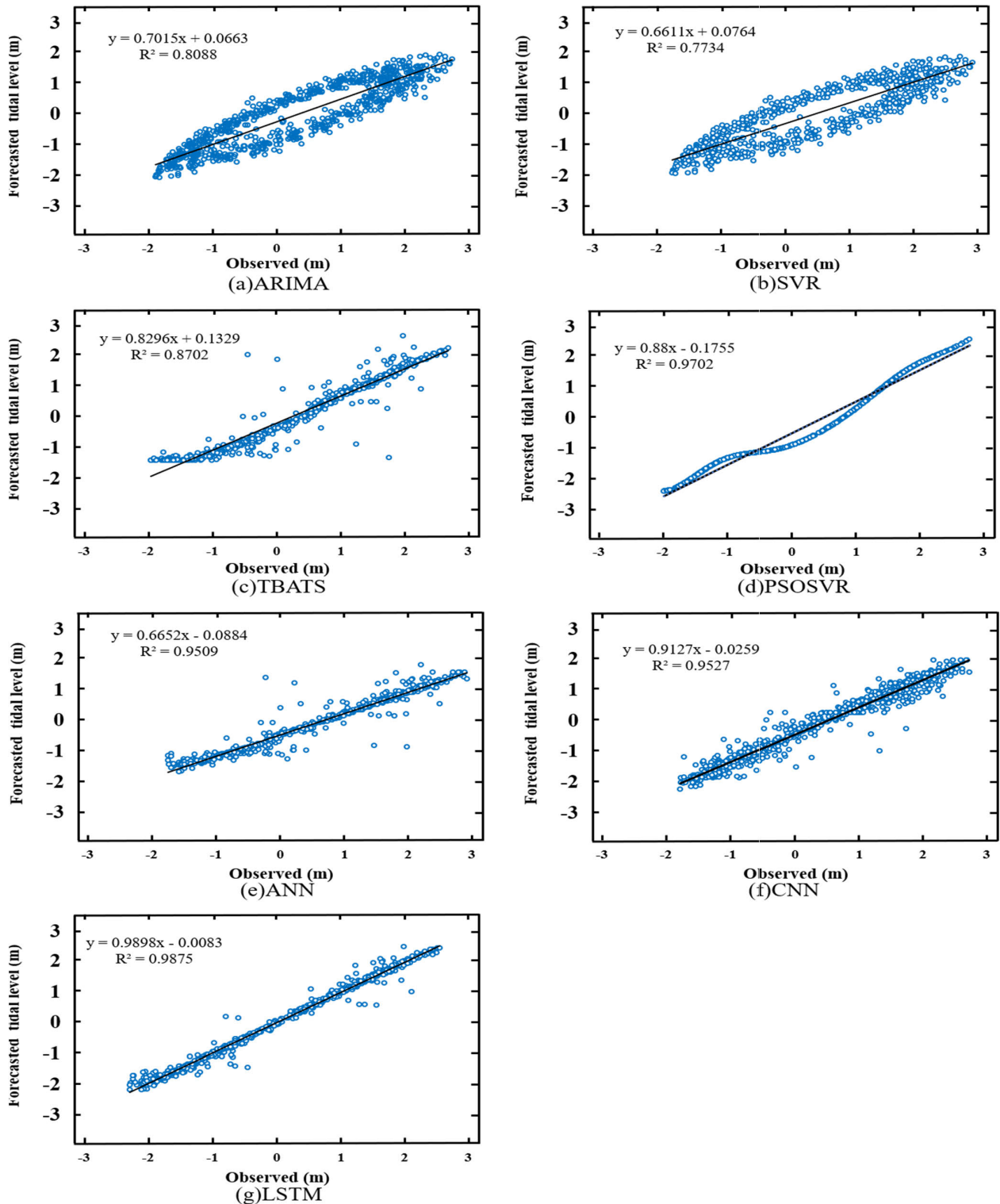


FIGURE 6. Scatterplots of observed versus forecasted tidal water levels from December 1 to December 30, 2018 (720 h prediction) for port 112 in different methods: (a) ARIMA, autoregressive integrated moving average; (b) SVR, support vector regression; (c) TBATS, Box-Cox transformation, ARMA errors, and trend seasonal components, (d) PSOSVR, particle swarm optimization-based support vector regression; (e) ANN, artificial neural network; (f) CNN, convolutional neural network; (g) LSTM, long short-term memory.

of the MAPE was 158%, 20%, 22%, and 26%; however, the error values of the ARIMA, ANN, and LSTM models increased slowly [Fig. 8(a), (e), and (g)]. The curve

was the most gentle, MAPE only increased by 10%, 9%, and 5% compared with the initial value. The LSTM model had the lowest rate of increase; the LSTM model's

TABLE 3. Tidal level forecasted from December 1 to December 30, 2018 by using different methods.

| Port | Criteria | ARIMA | TBATS | SVR | PSOSVR | ANN | CNN | LSTM |
|---------|----------------|----------|----------|----------|----------|----------|----------|--------------|
| 112 | MAPE(%) | 63.8 | 15.96 | 47.24 | 11.28 | 16.52 | 10.73 | 8.90 |
| | RMSE(m) | 0.40 | 0.24 | 0.22 | 0.13 | 0.11 | 0.06 | 0.06 |
| | R ² | 0.81 | 0.77 | 0.87 | 0.97 | 0.95 | 0.95 | 0.99 |
| 123 | MAPE(%) | 41.69 | 18.51 | 45.12 | 11.17 | 16.32 | 11.87 | 7.09 |
| | RMSE(m) | 0.55 | 0.16 | 0.20 | 0.12 | 0.13 | 0.06 | 0.01 |
| | R ² | 0.89 | 0.95 | 0.83 | 0.96 | 0.98 | 0.96 | 0.99 |
| 1102 | MAPE(%) | 60.14 | 19.32 | 47.19 | 14.3 | 11.38 | 10.73 | 9.70 |
| | RMSE(m) | 0.60 | 0.18 | 0.22 | 0.19 | 0.16 | 0.09 | 0.07 |
| | R ² | 0.73 | 0.85 | 0.80 | 0.92 | 0.95 | 0.96 | 0.99 |
| 1116 | MAPE(%) | 56.8 | 18.66 | 44.13 | 12.92 | 15.09 | 12.26 | 6.30 |
| | RMSE(m) | 0.44 | 0.28 | 0.18 | 0.09 | 0.16 | 0.08 | 0.06 |
| | R ² | 0.86 | 0.96 | 0.77 | 0.98 | 0.97 | 0.97 | 0.98 |
| 1166 | MAPE(%) | 61.92 | 18.3 | 42.29 | 10.08 | 11.04 | 11.17 | 7.23 |
| | RMSE(m) | 0.43 | 0.24 | 0.16 | 0.11 | 0.13 | 0.13 | 0.06 |
| | R ² | 0.82 | 0.96 | 0.90 | 0.94 | 0.96 | 0.94 | 0.96 |
| 1176 | MAPE(%) | 46.04 | 20.64 | 46.58 | 12.37 | 13.63 | 12.18 | 7.28 |
| | RMSE(m) | 0.56 | 0.18 | 0.21 | 0.11 | 0.15 | 0.11 | 0.06 |
| | R ² | 0.72 | 0.81 | 0.83 | 0.90 | 0.92 | 0.92 | 0.93 |
| 1246 | MAPE(%) | 64.12 | 21.72 | 41.3 | 13.26 | 12.55 | 11.13 | 7.45 |
| | RMSE(m) | 0.50 | 0.27 | 0.16 | 0.14 | 0.14 | 0.12 | 0.03 |
| | R ² | 0.09 | 0.92 | 0.90 | 0.93 | 0.95 | 0.96 | 0.96 |
| 1256 | MAPE(%) | 45.44 | 16.11 | 44.99 | 11.56 | 13.3 | 11.52 | 5.92 |
| | RMSE(m) | 0.53 | 0.28 | 0.20 | 0.09 | 0.12 | 0.10 | 0.03 |
| | R ² | 0.83 | 0.98 | 0.87 | 0.96 | 0.98 | 0.96 | 0.99 |
| 1276 | MAPE(%) | 63.3 | 21.02 | 46.66 | 11.31 | 12.74 | 12.91 | 6.50 |
| | RMSE(m) | 0.44 | 0.28 | 0.22 | 0.12 | 0.16 | 0.17 | 0.04 |
| | R ² | 0.82 | 0.9 | 0.89 | 0.94 | 0.94 | 0.95 | 0.96 |
| 1356 | MAPE(%) | 40.18 | 21.11 | 43.96 | 12.02 | 13.8 | 11.09 | 6.35 |
| | RMSE(m) | 0.43 | 0.27 | 0.15 | 0.12 | 0.15 | 0.11 | 0.06 |
| | R ² | 0.79 | 0.96 | 0.84 | 0.93 | 0.96 | 0.97 | 0.97 |
| 1366 | MAPE(%) | 52.48 | 21.76 | 45.37 | 11.68 | 11.97 | 13.07 | 6.74 |
| | RMSE(m) | 0.56 | 0.26 | 0.20 | 0.08 | 0.11 | 0.13 | 0.06 |
| | R ² | 0.78 | 0.97 | 0.86 | 0.95 | 0.98 | 0.96 | 0.98 |
| 1396 | MAPE(%) | 53.61 | 19.15 | 48.11 | 12.32 | 13.50 | 12.00 | 7.53 |
| | RMSE(m) | 0.49 | 0.23 | 0.24 | 0.11 | 0.15 | 0.13 | 0.04 |
| | R ² | 0.70 | 0.83 | 0.77 | 0.88 | 0.89 | 0.90 | 0.91 |
| 1436 | MAPE(%) | 51.36 | 16.36 | 47.15 | 12.68 | 11.92 | 13.95 | 7.14 |
| | RMSE(m) | 0.47 | 0.19 | 0.23 | 0.14 | 0.18 | 0.21 | 0.11 |
| | R ² | 0.87 | 0.96 | 0.87 | 0.96 | 0.95 | 0.96 | 0.97 |
| 1486 | MAPE(%) | 41.88 | 15.3 | 46.83 | 10.24 | 12.28 | 10.92 | 4.38 |
| | RMSE(m) | 0.54 | 0.25 | 0.22 | 0.08 | 0.17 | 0.09 | 0.03 |
| | R ² | 0.87 | 0.89 | 0.86 | 0.90 | 0.86 | 0.91 | 0.99 |
| 1496 | MAPE(%) | 42.16 | 17.03 | 44.19 | 12.68 | 14.02 | 11.37 | 7.88 |
| | RMSE(m) | 0.50 | 0.18 | 0.18 | 0.09 | 0.14 | 0.08 | 0.02 |
| | R ² | 0.69 | 0.92 | 0.78 | 0.95 | 0.95 | 0.96 | 0.99 |
| 1516 | MAPE(%) | 48.87 | 17.75 | 43.12 | 12.46 | 12.94 | 10.34 | 8.33 |
| | RMSE(m) | 0.58 | 0.28 | 0.18 | 0.10 | 0.12 | 0.12 | 0.10 |
| | R ² | 0.85 | 0.92 | 0.86 | 0.96 | 0.97 | 0.95 | 0.98 |
| 1586 | MAPE(%) | 52.89 | 19.86 | 44.56 | 13.05 | 16.08 | 12.90 | 4.23 |
| | RMSE(m) | 0.50 | 0.22 | 0.18 | 0.11 | 0.17 | 0.09 | 0.02 |
| | R ² | 0.83 | 0.86 | 0.88 | 0.98 | 0.91 | 0.98 | 1.00 |
| Average | MAPE(%) | 52.15*** | 18.73*** | 45.22*** | 12.08*** | 13.47*** | 11.77*** | 6.997 |
| | RMSE(m) | 0.50*** | 0.23*** | 0.19*** | 0.11*** | 0.14*** | 0.10*** | 0.049 |

MAPE, mean absolute percentage error; RMSE, root mean square error; R², coefficient of determination; boldface, the optimal values in each row. ARIMA, autoregressive integrated moving average; SVR, support vector regression; TBATS, trigonometric seasonality, Box-Cox transformation, ARMA errors, and trend seasonal components; PSOSVR, particle swarm optimization-based support vector regression; ANN, artificial neural network; CNN, convolutional neural network; LSTM, long short-term memory. * $p < .05$, ** $p < .01$, *** $p < .001$.

long-term prediction performance was thus excellent and stable. Figs. 9 and 10 illustrate the MAPE and RMSE of the seven methods in each port. The LSTM model consistently

had the lowest RMSE and MAPE. Therefore, the LSTM method was more stable and accurate than were the other six methods.

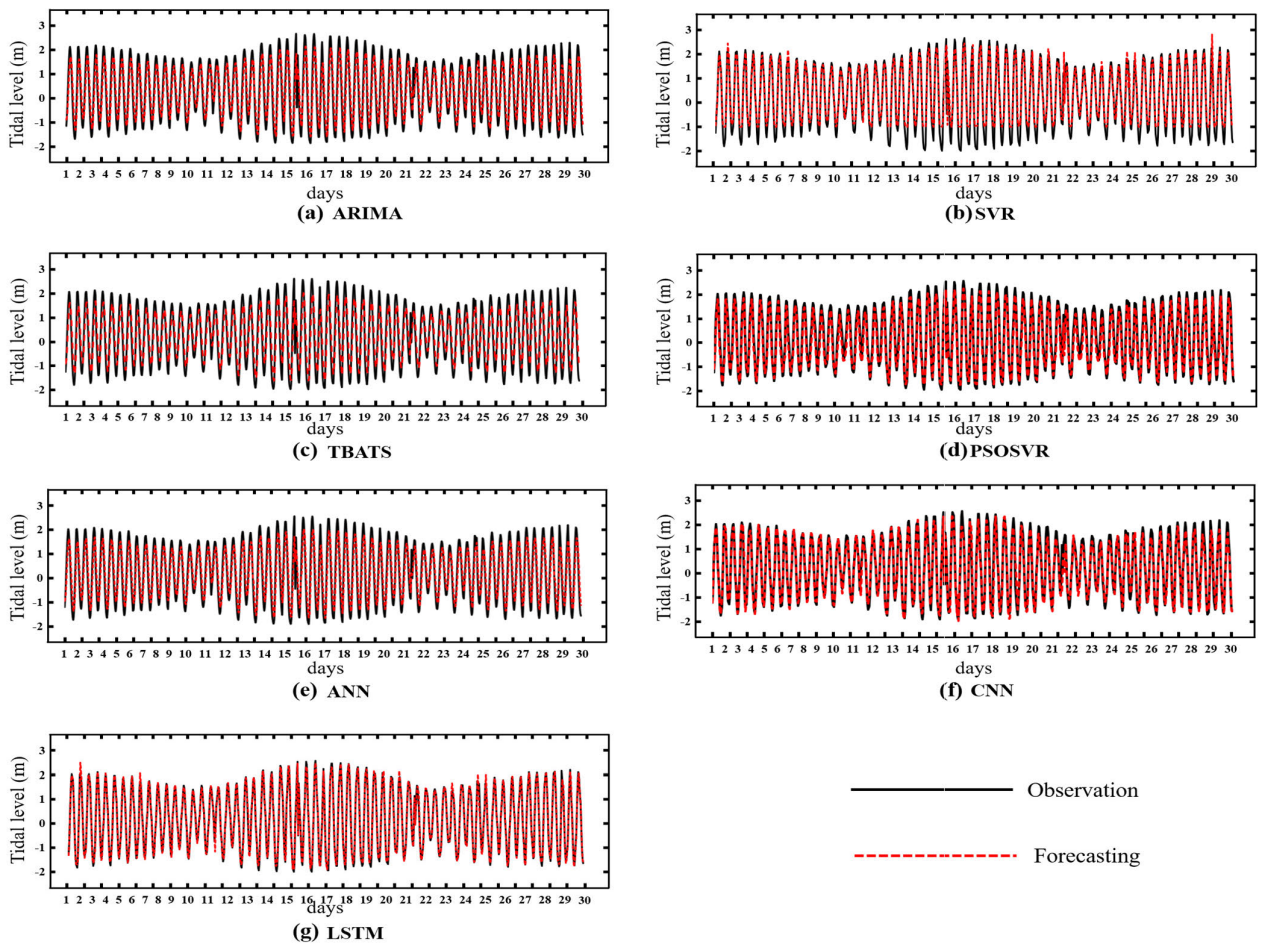


FIGURE 7. Tidal water levels forecasting result for port 112 when using seven methods for different forecast lengths: (a) ARIMA, (b) SVR, (c) TBATS, (d) PSOSVR, (e) ANN, (f) CNN, and (g) LSTM models.

To verify the advantages and disadvantages of the proposed LSTM forecasting model, a Wilcoxon signed-rank test was conducted to compare the MAPE and RMSE values of the ARIMA, SVR, TBATS, PSOSVR, ANN, CNN, and LSTM methods. To analyze the differences in the prediction results of each method, a null hypothesis (H_0) and alternative hypothesis (H_1) were considered. H_0 posited that the LSTM method's prediction results were not significantly different from those of the other six methods, and H_1 posited the converse (that the prediction results significantly differed). The prediction results differed with statistical significance if $p < .05$. As detailed in Table 3, the LSTM method had a significantly lower RMSE and MAPE than did the other six methods; thus, the LSTM method outperformed the other methods in tidal level prediction. According to the comparison test results, the LSTM method was more likely than the other methods to provide a robust forecasting model with a small error rate.

IV. DISCUSSION

A. COMPARISON OF ARIMA AND TBATS

The ARIMA model, with autoregressive movement assumed, must refer to a large quantity of historical data to determine

the optimal parameter combination of the model. TBATS was proposed in 2011. Statistical hybrid models were designed to compensate for inaccurate prediction results when a single model is used: it pools the advantages of multiple models and increases the number of calculations. However, it has the disadvantage of being computationally expensive. Moreover, and in general, statistical models still solve nonlinear problems with difficulty. In Fig 9 and 10, ARIMA and TBATS consistently had the highest error values.

B. COMPARISON OF SVR AND PSOSVR

Machine learning has been popular and widely applied. Prior to Hinton's pioneering formulation of machine learning for the time series prediction problem, SVR was the dominant machine learning method. Therefore, we selected SVR as a representative method for comparison.

The use of a single machine learning method is unlikely to yield the best accuracy, and hybrid methods are likely to be better. For example, PSOSVR uses PSO optimization algorithms to improve on the shortcomings of the original SVR. PSO, developed by Eberhart and Kennedy, is a population-based iterative optimization algorithm inspired by the social behavior of bird flocking. Because SVR has

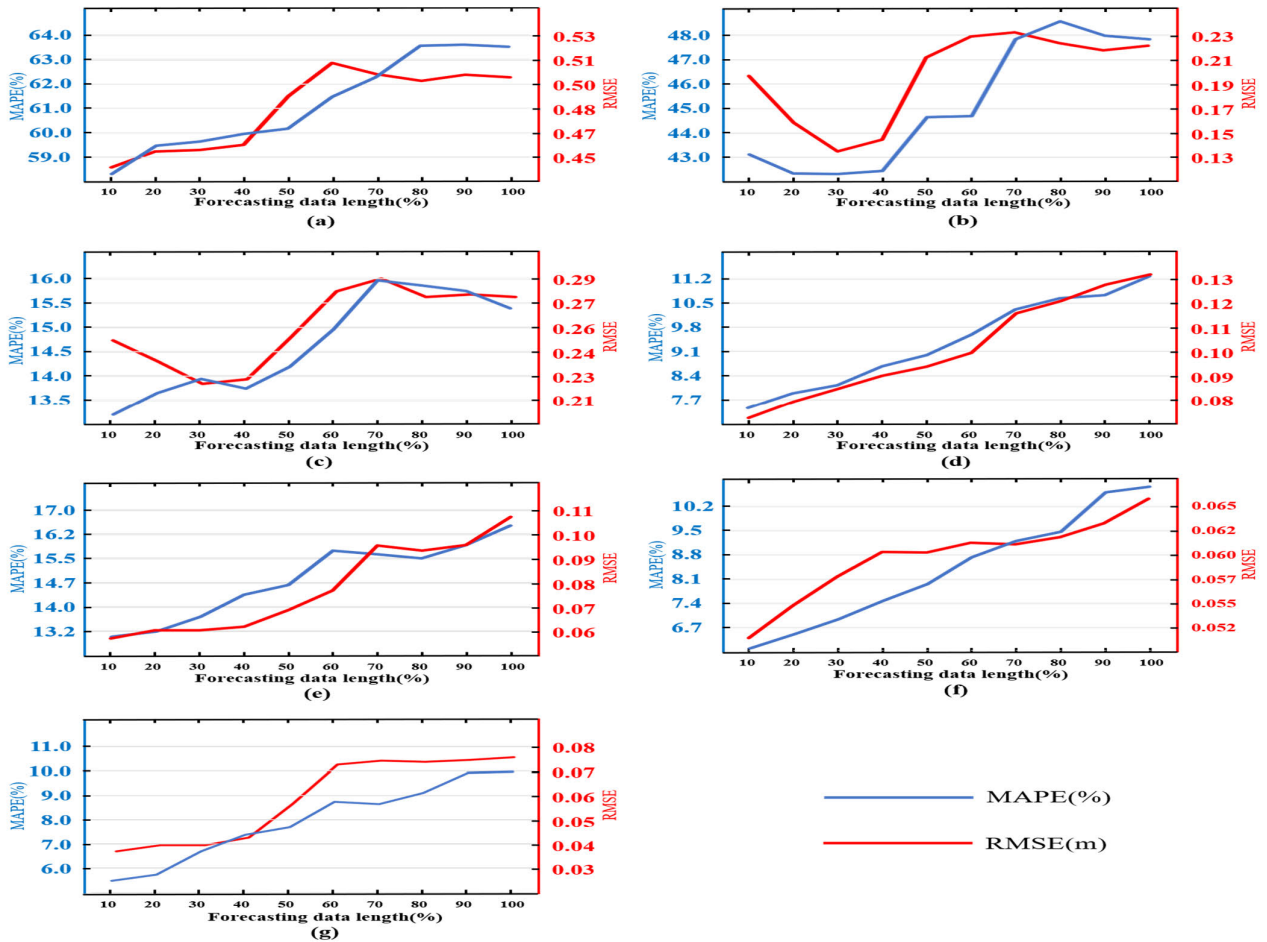


FIGURE 8. Tidal water level forecasting result for port 112 when using different forecast lengths and methods: (a) ARIMA, (b) SVR, (c) TBATS, (d) PSOSVR, (e) ANN, (f) CNN, and (g) LSTM.

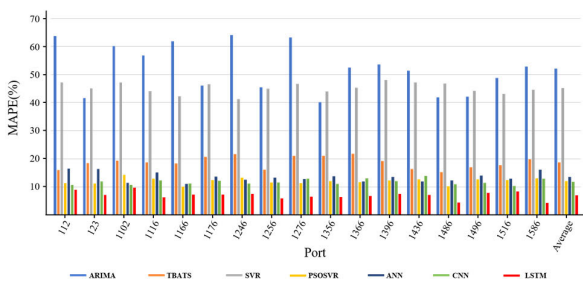


FIGURE 9. MAPE of the six methods for different forecast lengths.

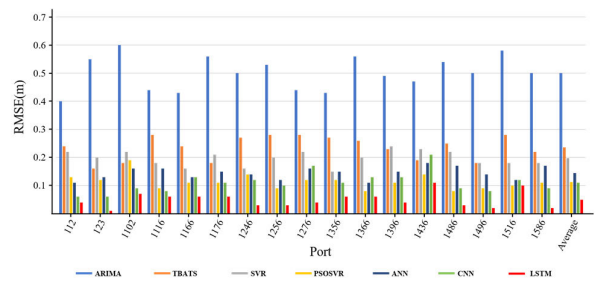


FIGURE 10. RMSE of the six methods for different forecast lengths.

three hyperparameters—the regularization parameter (C), bandwidth of the kernel function (σ), and tube size of the ε -insensitive loss function (ε)—differences among these parameters greatly affect SVR’s forecasting accuracy. The automatic adjustment of these three hyperparameters in SVR remains a prominent challenge for improving SVR’s forecasting accuracy. Through an optimized algorithm, PSO can help to adjust the hyperparameters to their appropriately selected values to avoid either overfitting or underfitting [38]. This study proved that the hyperparameters is crucial for SVR. Therefore, compared with SVR, PSOSVR has a lower MAPE and RMSE (Figs. 9 and 10).

C. COMPARISON OF ANN, CNN AND LSTM

ANNs is a part of a neural network, which simulates how the human brain learns. ANN is a model constructed by transmitting neurons, and it has the advantage of good feature extraction capabilities. After the backpropagation neural network was proposed, ANNs also allowed shallow neural networks to contribute distinctively in machine learning. Conventional ANNs with shallow architectures are difficult to train if they become too complex—for example, when the network includes many layers and, consequently,

many parameters. It has been widely demonstrated that deep ANN architectures, known as deep neural networks (DNNs), outperform conventional shallow ANN architectures in several applications [48]. Recently, deep learning has gained substantial popularity in the machine learning community because it is considered a general framework that facilitates the training of deep neural networks with many hidden layers [49]. Many neural network models have been widely used to solve several types of time series forecasting problems. Of these models, recurrent neural networks (RNNs) have received much attention [48], [50].

The reason for such attention is that RNNs are a class of ANN models that possess an internal state or short-term memory due to recurrent feedback connections, making RNNs suitable for modeling sequential or time series data. In such modeling, the RNN maintains a vector of activation parameters for each time step, especially when short-term dependencies are included in the input data. However, if an RNN is trained using stochastic gradient descent, the RNN has difficulty learning long-term dependencies that are encoded in the input sequences due to the vanishing gradient problem [51], [52]. To then allow the LSTM to learn long-term dependencies, a specialized neuron or cell structure is employed in the LSTM network, which maintains constant backward flow in the error signal [24], [53].

Recently, deep learning in general and CNN in particular has become the methodology of choice for image analysis, following its tremendous success in routine computer vision applications [54], [55]. CNN is more data efficient than fully connected networks are due to CNN's translation weight sharing properties in the convolutional layers. Layers in CNN are translation equivariant (i.e., when the network input is shifted, internal representations are also shifted), which makes transnational weight sharing effective in each layer. Time series with trends are the most common data sets used in forecasting. Both the convolutional layer and the pooling layer of a CNN can be used to extract crucial features and patterns that reflect the seasonality, trends, and time lag correlation coefficients in the data. Therefore, in addition to being applied in image classification [56], semantic segmentation [57], and object detection [58], CNN is applied in time series forecasting.

Compared with CNNs and LSTMs, ANNs have the advantages of lower computational cost and fewer neurons and hidden layers but the disadvantage of a higher rate of error. Although CNN has strong feature extraction capabilities, it is mostly used in tasks such as image classification or image recognition. Because CNN could not leverage on its advantages in this study, its error rate was similar to that of ANN. Because the intervals between past and future timepoints are crucial to solving time series problems, LSTM, by virtue of its special gate structure, can determine the relationship between past and future timepoints and thus furnish accurate predictions in a long series.

V. CONCLUSION

In this study, tidal water level prediction indicators and related time series forecasting models were examined to increase the transportation efficiency and safety of entry for large vessels in commercial ports. The LSTM method was proposed for long-term tidal water level prediction. The proposed method was tested using 16 datasets. The performance of the proposed method was compared with that of five time series forecasting models, namely the ARIMA, SVR, TBATS, PSOSVR, ANN, and CNN models. The experimental results indicated that the LSTM model had a lower MAPE and RMSE than the other six forecasting models did. The LSTM model also had the highest R^2 value among all the compared models in all the prediction intervals. Thus, the LSTM forecasting model provided more robust results than the other five models did. In addition to lower error values, the LSTM model exhibited a higher stability and more rapid convergence compared with the other five models during training. The prediction results of the LSTM model were smooth and interpretable. The LSTM model also exhibited a superior fit with the collected data compared with the other models. Forecasting the tidal water level with a higher accuracy than that achieved in this study may result in a considerable increase in training time. Therefore, in future studies, a suitable method should be developed for achieving increased accuracy without increasing the training dimensions and time.

REFERENCES

- [1] A. Becker, S. Inoue, M. Fischer, and B. Schwegler, "Climate change impacts on international seaports: Knowledge, perceptions, and planning efforts among port administrators," *Climatic Change*, vol. 110, nos. 1–2, pp. 5–29, Jan. 2012, doi: [10.1007/s10584-011-0043-7](https://doi.org/10.1007/s10584-011-0043-7).
- [2] Taiwan International Ports Corporation. (May 22, 2019). *Annual Statistical Reports*. [Online]. Available: <https://www.twport.com.tw/en/cp.aspx?n=8E9F2CDAA08A72A7>
- [3] G. Godin, "The analysis of tides," Univ. Toronto Press, Toronto, ON, Canada, Tech. Rep. 551.4708 G6, 1972, p. 264.
- [4] S. Li, L. Liu, S. Cai, and G. Wang, "Tidal harmonic analysis and prediction with least-squares estimation and inaction method," *Estuarine, Coastal Shelf Sci.*, vol. 220, pp. 196–208, May 2019, doi: [10.1016/j.ecss.2019.02.047](https://doi.org/10.1016/j.ecss.2019.02.047).
- [5] W. H. Munk and D. E. Cartwright, "Tidal spectroscopy and prediction," *Philos. Trans. Roy. Soc. London. A, Math. Phys. Sci.*, vol. 259, no. 1105, pp. 533–581, 1966.
- [6] E. P. Flinchem and D. A. Jay, "An introduction to wavelet transform tidal analysis methods," *Estuarine Coast Shelf Sci.*, vol. 51, no. 2, pp. 177–200, Aug. 2000, doi: [DOI.10.1006/ecss.2000.0586](https://doi.org/10.1006/ecss.2000.0586).
- [7] C. P. Tsai and T. L. Lee, "Back-propagation neural network in tidal-level forecasting," *J. Waterway, Port Coastal, Ocean Eng.*, vol. 125, no. 4, pp. 195–202, 1999, doi: [Doi.10.1061/\(ASCE\)0733-950x\(1999\)125:4\(195\)](https://doi.org/10.1061/(ASCE)0733-950x(1999)125:4(195)).
- [8] O. Sahin, R. A. Stewart, G. Faivre, D. Ware, R. Tomlinson, and B. Mackey, "Spatial Bayesian network for predicting sea level rise induced coastal erosion in a small pacific island," *J. Environ. Manage.*, vol. 238, pp. 341–351, May 2019, doi: [10.1016/j.jenvman.2019.03.008](https://doi.org/10.1016/j.jenvman.2019.03.008).
- [9] M. Minsky and S. Papert, *Perceptron: An Introduction to Computational Geometry*, vol. 19, no. 88. Cambridge, MA, USA: MIT Press, 1969, p. 2.
- [10] F. P. Bretherton, R. E. Davis, and C. Fandry, "A technique for objective analysis and design of oceanographic experiments applied to MODE-73," in *Deep Sea Research and Oceanographic Abstracts*. vol. 23, no. 7. Amsterdam, The Netherlands: Elsevier, 1976, pp. 559–582.
- [11] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8599–8603.

- [12] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [14] C.-H. Yang, S.-H. Moi, F. Ou-Yang, L.-Y. Chuang, M.-F. Hou, and Y.-D. Lin, "Identifying risk stratification associated with a cancer for overall survival by deep learning-based CoxPH," *IEEE Access*, vol. 7, pp. 67708–67717, 2019, doi: [10.1109/Access.2019.2916586](https://doi.org/10.1109/Access.2019.2916586).
- [15] Z. Y. Tang, C. Dealmeida, and P. A. Fishwick, "Time series forecasting using neural networks vs. Box-Jenkins methodology," *Simulation*, vol. 57, no. 5, pp. 303–310, 1991, doi: [Doi.10.1177/003754979105700508](https://doi.org/10.1177/003754979105700508).
- [16] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," 2019, *arXiv:1909.00590*. [Online]. Available: <http://arxiv.org/abs/1909.00590>
- [17] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112896, doi: [10.1016/j.eswa.2019.112896](https://doi.org/10.1016/j.eswa.2019.112896).
- [18] K. E. Leffler and D. A. Jay, "Enhancing tidal harmonic analysis: Robust (hybrid) solutions," *Continental Shelf Res.*, vol. 29, no. 1, pp. 78–88, Jan. 2009.
- [19] M.-Y. Lin, W.-Y. Sun, M.-D. Chiou, C.-Y. Chen, H.-Y. Cheng, and C.-H. Chen, "Development and evaluation of a storm surge warning system in Taiwan," *Ocean Dyn.*, vol. 68, no. 8, pp. 1025–1049, Aug. 2018.
- [20] K. Yin, S. Xu, W. Huang, and Y. Xie, "Effects of sea level rise and typhoon intensity on storm surge and waves in pearl river estuary," *Ocean Eng.*, vol. 136, pp. 80–93, May 2017.
- [21] T.-L. Lee, "Back-propagation neural network for long-term tidal predictions," *Ocean Eng.*, vol. 31, no. 2, pp. 225–238, Feb. 2004.
- [22] A. Riazi, "Accurate tide level estimation: A deep learning approach," *Ocean Eng.*, vol. 198, Feb. 2020, Art. no. 107013.
- [23] H. H. H. Aly and M. E. El-Hawary, "A proposed ANN and FLSM hybrid model for tidal current magnitude and direction forecasting," *IEEE J. Ocean. Eng.*, vol. 39, no. 1, pp. 26–31, Jan. 2014, doi: [10.1109/Joe.2013.2241934](https://doi.org/10.1109/Joe.2013.2241934).
- [24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: [DOI.10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [25] J. Fagerström, M. Bång, D. Wilhelms, and M. S. Chew, "LiSep LSTM: A machine learning algorithm for early detection of septic shock," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019.
- [26] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1764–1772.
- [27] K.-Y. Huang, C.-H. Wu, and M.-H. Su, "Attention-based convolutional neural network and long short-term memory for short-term detection of mood disorders based on elicited speech responses," *Pattern Recognit.*, vol. 88, pp. 668–678, Apr. 2019.
- [28] P. Nejedly, V. Kremen, V. Sladky, J. Cimbalka, P. Klimes, F. Plesinger, I. Viscor, M. Pail, J. Halamek, B. H. Brinkmann, M. Brazdil, P. Jurak, and G. Worrell, "Exploiting graphoelements and convolutional neural networks with long short term memory for classification of the human electroencephalogram," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.
- [29] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994, doi: [10.1109/72.279181](https://doi.org/10.1109/72.279181).
- [30] A. Graves, "Generating sequences with recurrent neural networks," 2013, *arXiv:1308.0850*. [Online]. Available: <http://arxiv.org/abs/1308.0850>
- [31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [32] G. E. Box and G. M. Jenkins, "Time series analysis: Forecasting and control San Francisco," Holden-Day, San Francisco, CA, USA, Tech. Rep. 9780816211043, 1976, vol. 13.
- [33] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *J. Amer. Stat. Assoc.*, vol. 106, no. 496, pp. 1513–1527, Dec. 2011, doi: [10.1198/jasa.2011.tm09771](https://doi.org/10.1198/jasa.2011.tm09771).
- [34] V. Vapnik, S. E. Golowich, and A. J. Smola, "Support vector method for function approximation, regression estimation and signal processing," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 281–287.
- [35] V. Valdimir and N. Vapnik, *The Nature of Statistical Learning Theory*. Cham, Switzerland: Springer, 1995.
- [36] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999, doi: [10.1109/72.788640](https://doi.org/10.1109/72.788640).
- [37] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE ICNN*, vol. 4, Nov./Dec. 1995, pp. 1942–1948.
- [38] H. H. Liu, L. C. Chang, C. W. Li, and C. H. Yang, "Particle swarm optimization-based support vector regression for tourist arrivals forecasting," *Comput. Intell. Neurosci.*, vol. 2018, Sep. 2018, Art. no. 6076475, doi: [Artn.607647510.1155/2018/6076475](https://doi.org/10.1155/2018/6076475).
- [39] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
- [40] A. Cauchy, "Méthode générale pour la résolution des systèmes d'équations simultanées," *Comp. Rend. Sci. Paris*, vol. 25, no. 1847, pp. 536–538, 1847.
- [41] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989, doi: [DOI.10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541).
- [42] J.-S. Chen and P.-C. Lin, "An intelligent financial ratio selection mechanism for earning forecast (special issue on theory, methodology and applications in financial engineering)," *J. Oper. Res. Soc. Jpn.*, vol. 45, no. 4, pp. 373–384, 2002.
- [43] S. Lawrence, C. L. Giles, A. Chung Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.
- [44] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2015.
- [45] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013, doi: [10.1109/TPAMI.2012.59](https://doi.org/10.1109/TPAMI.2012.59).
- [46] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, Feb. 2017.
- [47] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised learning," *Int. J. Comput. Sci.*, vol. 1, no. 2, pp. 111–117, 2006.
- [48] A. Sagheer and M. Kothb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019, doi: [10.1016/j.neucom.2018.09.082](https://doi.org/10.1016/j.neucom.2018.09.082).
- [49] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015, doi: [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003).
- [50] A. Tealab, "Time series forecasting using artificial neural networks methodologies: A systematic review," *Future Comput. Informat. J.*, vol. 3, no. 2, pp. 334–340, Dec. 2018.
- [51] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen Netzen," (in German), Diploma thesis, Technische Univ. München, Munich, Germany, 1991.
- [52] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Networks*. Piscataway, NJ, USA: IEEE Press, 2001.
- [53] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000, doi: [Doi.10.1162/089976600300015015](https://doi.org/10.1162/089976600300015015).
- [54] M. I. Razzak and S. Naz, "Microscopic blood smear segmentation and classification using deep contour aware CNN and extreme machine learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 49–55.
- [55] M. I. Razzak and B. Alhaqabani, "Automatic detection of malarial parasite using microscopic blood images," *J. Med. Imag. Health Informat.*, vol. 5, no. 3, pp. 591–598, Jun. 2015, doi: [10.1166/jmihi.2015.1417](https://doi.org/10.1166/jmihi.2015.1417).
- [56] S. Kumar Roy, S. Manna, S. Ram Dubey, and B. Baran Chaudhuri, "LiSHT: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks," 2019, *arXiv:1901.05894*. [Online]. Available: <http://arxiv.org/abs/1901.05894>

[57] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[58] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.



CHENG-HONG YANG (Senior Member, IEEE) received the Ph.D. degree in computer engineering from North Dakota State University, in 1992. He has served as the President of the National Kaohsiung University of Applied Science, Taiwan, from 2012 to 2016. He is currently the Chair Professor with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan. He has authored/coauthored over 380 refereed publications and a number of book chapters. His research interests include fuzzy control, evolutionary computation, bioinformatics, data analysis, and the applications of these methods. He is a Fellow of the Institution of Engineering and Technology, and the American Biographical Institute. He is an Editorial Board Member of multiple international journals.



CHIH-HSIEN WU received the B.S. degree from the Department of Information Management, Shu-Te University, in 2019. He is currently pursuing the M.S. degree with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan. He has rich experience in machine learning, computer programming, database design and management, and systems programming and design. His research interests include forecasting tide level and deep learning.



CHIH-MIN HSIEH was born in Taiwan, in 1973. He received the master's and Ph.D. degrees in ocean engineering from the Department of Hydraulic and Ocean Engineering, National Cheng Kung University. From 2006 to 2007, he was a Postdoctoral Fellow with the Institute of Physics, Academia Sinica, Taipei City, Taiwan. He is currently a Professor with the Department of Maritime Information and Technology, National Kaohsiung University of Science and Technology. His research interests include the signal analysis and computed fluid dynamics for wave dynamics, sediment transport, and coastal engineering.

...