ORIGINAL RESEARCH

Multi-step Ahead Urban Water Demand Forecasting Using Deep Learning Models

Bibhuti Bhusan Sahoo[1] · Banamali Panigrahi[2] · Trushnamayee Nanda[3] · Mukesh Kumar Tiwari[4] · Sovan Sankalp[5]

Abstract

Accurate prediction of water demand in a city is crucial for the management of urban water distribution system. The current study aims to create adequate daily water demand forecasting models for the Canadian metropolis of London utilizing deep learning (DL)-based models. This study explores the potential of two stand-alone DL models for daily water demand modeling using a convolutional neural network (CNN) and the long short-term memory (LSTM) along with their hybrid CNN–LSTM model. Furthermore, a deep learning-based bi-directional LSTM model is introduced with CNN to predict daily water consumption in London as CNN–BiLSTM hybrid model. Daily water consumption data for the years 2009 to 2020 are used for the development and assessment of the predictive models. These stand-alone and hybrid models have been developed with specified input lags of daily water demand and verified for daily water demand prediction. The performance of the developed hybrid models was compared with other well-established DL-based stand-alone models. The model outcomes during the training, validation, and testing phases were assessed using statistical metrics such as the mean absolute error (MAE), Nash–Sutcliffe coefficient (NSE), correlation coefficient (r), Scatter Index (SI), Mean Bias Error (MBE), and Discrepancy Ratio (DR). The stand-alone models captured the observations very well during training and testing which is obvious at 1-day ahead. Moreover, at 7 days and 15 days ahead those models, except the stand-alone CNN, closely reproduced the pattern in daily water demand. Among the hybrid models, the outperforming CNN–BiLSTM model produced 1-day to 15-day multi-step ahead forecasting with performance metrics in the following ranges: MAE = 0.245–2.541 ML/day, NSE = 99.830–84.843% and r = 0.999–0.921 during the testing period. The uncertainty analysis has been performed which advocates the superiority of the CNN–BiLSTM model showing the forecasts bands with 88–90% observations within the 90% confidence interval (CI). Overall, the outcomes supporting the CNN–BiLSTM is to be considered as a promising deep learning procedure for accurate forecasting of urban water demand in any city globally.

Keywords Water demand · Deep learning · Urban water · Neural network models · Forecasting

✉ Banamali Panigrahi
  bap654@usask.ca

  Bibhuti Bhusan Sahoo
  bibhuti5000@gmail.com

  Trushnamayee Nanda
  nanda.trushnamayee@gmail.com

  Mukesh Kumar Tiwari
  tiwari.iitkgp@gmail.com

  Sovan Sankalp
  sovan.sankalp@cutm.ac.in

[1] Department of Agricultural Engineering, Centurion University, Paralakhemundi, Odisha, India

[2] Toxicology Center, University of Saskatchewan, 44 Campus Drive, Saskatoon, SK S7N 5B3, Canada

[3] Soil and Water Conservation, Directorate of Soil Conservation and Watershed Development, Bhubaneswar, Odisha 768004, India

[4] College of Agricultural Engineering and Technology, Anand Agricultural University, Godhra 389001, India

[5] Department of Civil Engineering, Centurion University of Technology and Management, Bhubaneswar, Odisha 761211, India

Introduction

In any water distribution system water demand can be represented as the driver, and an effective water demand forecasting aids to accomplish effective water management in any urban metropolis. Accurate water demand forecasting helps in better urban planning and sustainable development [1, 2]. Water demand in a city is influenced by many factors such as the population of the city, various industrial establishments in the city, climatic conditions [3, 4]. Therefore, predicting accurate water demand in an urban city is a challenging task. Water demand forecasting can be classified into three types based on the time prospect considered: short-term, medium-term, and long-term forecasting [5]. The medium- and long-term water demand forecasting may assist in planning and development of the reservoir systems and water network [3, 6]. Short-term water demand forecasting also called as operational time forecast provides an estimate of water demand over short horizons ranging from few days to weeks. Short-term demand forecasts enable real-time management of water distribution network devices.

Traditionally, researchers used various techniques to model water demand such as regression, autoregressive integrated moving averages (ARIMA), artificial intelligence (AI) models [7–10]. Conventional techniques, particularly linear regression-based models are known for their simplicity in understanding and implementation. They have been the primary focus of research and extensively applied in practical applications [11]. On the other hand, they appear to lack accuracy in forecasting water demand, which can create serious problems in the smooth running of the network system. Moreover, changes in water demand exhibit nonlinearity due to non-linear changes in water consumption, temperature variations, and holidays impacting urban water usage [12] and cannot be accurately predicted by linear approaches [13]. In the realm of water demand forecasting, some of the AI models have been explored in previous studies such as artificial neural network (ANN) [5, 6], support vector machines (SVM) [13–16], SVM method using Fourier method [14], random forests [15], extreme learning machines (ELM) [15, 17], employed ELM in conjunction with wavelet based ANN [18], system dynamics modeling (SDM) [19], ensemble wavelet–bootstrap machine-learning approach [20], singular spectrum analysis coupled with neural networks [21], and adaptive neuro-fuzzy inference system (ANFIS) [22]. In general, these learning techniques have shown successful results and are widely applied in water demand forecasting.

However, the growth of urbanization and the consequences of climate change causes considerable uncertainty in the time-series data making forecasting more difficult, which inspiring experts to improve and develop new models including the use of deep learning (DL) techniques. In contrast to the traditional method for water demand forecasting, DL can take previous data and automatically recognize trends through consecutive layer representation, and layers leading to the development of the model which is also known as hierarchical representation learning [23]. In recent years, DL techniques have shown remarkable and promising results in many fields such as sentiment analysis [24, 25], facial recognition [26, 27], natural language processing [28], weather prediction [29], and time-series analysis [30]. A DL network allows the usage of hundreds or even thousands of hierarchical layers, and it learns autonomously from past data [31]. CNN network can very well extract the features in urban water consumption dataset while being used in urban water demand forecasting problem [32].

The incredible achievements of DL modes in various fields have inspired researchers to investigate and implement the DL model for forecasting water demand. Reviewing subsequent and more recent work on deep learning for water demand forecasting, we found that Guo G et al. [33] used the gated recurrent unit network (GRUN) model to forecast short-term water demand and Salloom T [34] used the GRUN for real-time water demand forecast in China. Study also used long short-term memory (LSTM) neural networks combined with wavelet transform and principal component analysis for daily urban water demand prediction [35] and the deep belief network (DBN) technique was used for modeling daily urban water demand [36]. Moreover, researchers applied a continuous deep belief echo state network (CDBESN) for hourly urban water demand forecasting and found the developed model outperforms the echo state network (ESN) and SVR models used in the study [37]. The recent study also addressed the challenge of predicting household water consumption for four different water use types (apartment, restaurant, detached house, and elementary school) with non-linear patterns (weekdays and weekends as explanatory variables) using deep learning-based LSTM models [38]. Furthermore, the study by Kavya M [39] focused on short-term water demand forecasting using nine machine learning and deep learning models. The results indicate that deep learning models, particularly the LSTM model, outperform machine learning models in both univariate and multivariate scenarios. The outcomes were compared with the artificial neural network (ANN) model, support vector regression (SVR) model, and conventional autoregressive integrated moving average (ARIMA) model. Researchers stated that the LSTM gained huge attraction in the field of time-series forecasting, particularly bi-directional LSTM (BiLSTM) introduced an architecture for better learning [40] and the prediction error is reduced using the hybrid

CNN–BiLSTM model [12]. Moreover, the hybrid model can incorporate historical water data and climatic factors, resulting in improved prediction accuracy compared to other stand-alone models (such as LSTM, BiLSTM, CNN, GRUN, and ANN). In addition, it demonstrates shorter training time and convergence in comparison to other models [12].

The objectives in the present study are: (i) to evaluate the predictive efficacy of DL models, namely CNN, LSTM, BiL-STM models and (ii) to develop hybrid models CNN–LSTM and CNN–BiLSTM that combines the advantages of stand-alone models (CNN and LSTM) and bidirectional approach, respectively. To the best of our knowledge, the potential of the hybrid CNN–LSTM and CNN–BiLSTM deep learning model for water demand prediction has not been explored yet. Further, the uncertainty analysis has been done for all the models used in this study for better evaluation of model limitations.

Study Area and Data Collection

The daily water consumption datasets for the London city in Canada (study area located in Fig. 1) have been used from 1st July 2009 to 2nd September 2020 for development of urban water demand forecasting models. The daily water demand data are computed from water billing data and provided by the Municipal Artificial Intelligence application lab out of the Information Technology Services division (https://github.com/aildnont/water-forecast). The variation of the daily water consumption in the city can be observed from Fig. 2 which reaches a peak between March and November with few exceptions during the year 2014 and beginning of 2017. The descriptive statistics of the water demand data from the city of London, including minimum, maximum, mean, and standard deviation can be found in Table 1.

Methodology

This section describes the state-of-the-art deep learning methods which have become the proposed solutions for time-series forecasting in water resources variables related to water quantity.

Neural Network and Its Forms

The state-of-the art deep learning models have been evolved from the age-old artificial neural network [41] with different levels of complexity in structure, parameterization and run-time consumption. The models are subsequently evaluated for 1 day to 15 days ahead of water demand forecasting. The artificial neural networks are based on learning the non-linear input–output relationship existing in the observed parameters through weighted synaptic connections [42] consisting of input layer, one or more processing layers, and an output layer represented by nodes/neurons The model performance depends on historical data used for training and the method for determining the weights and functions for inputs and nodes during training [43]. The neuron has an internal activation level in the hidden layer called the activation function/transfer function which establishes a relationship between the weighted inputs and the outputs.

Furthermore, the structure of recurrent neural network (RNN) is meant to handle sequential data sample or ordered data in which subsequent things relate and follow each other. RNN is having an internal state (memory) that allows accepting the short- or long-term outputs to be used as inputs [44]. RNNs can process input data sequences using their internal hidden memory, and as a result, they can be used for applications such as handwriting identification, speech recognition, and time-series analysis. In RNN, the information cycles through a loop make the inputs related to each other. This helps RNN to examine the input and
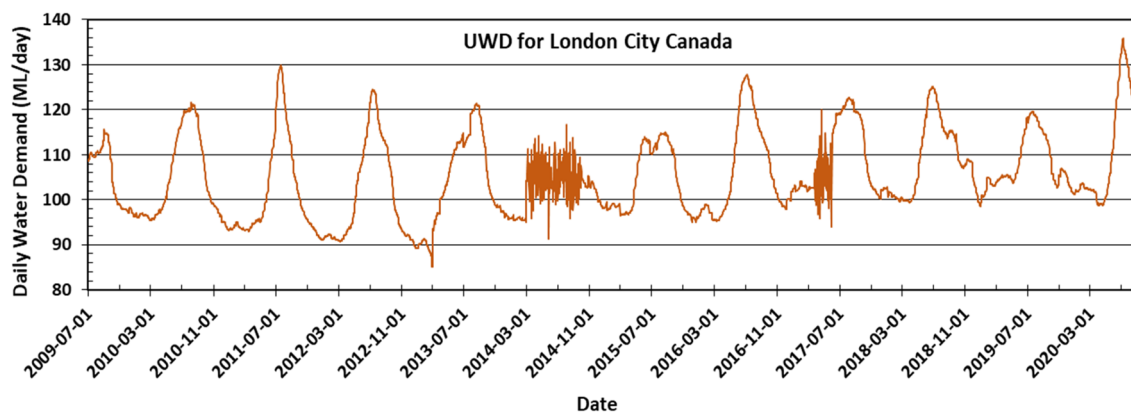

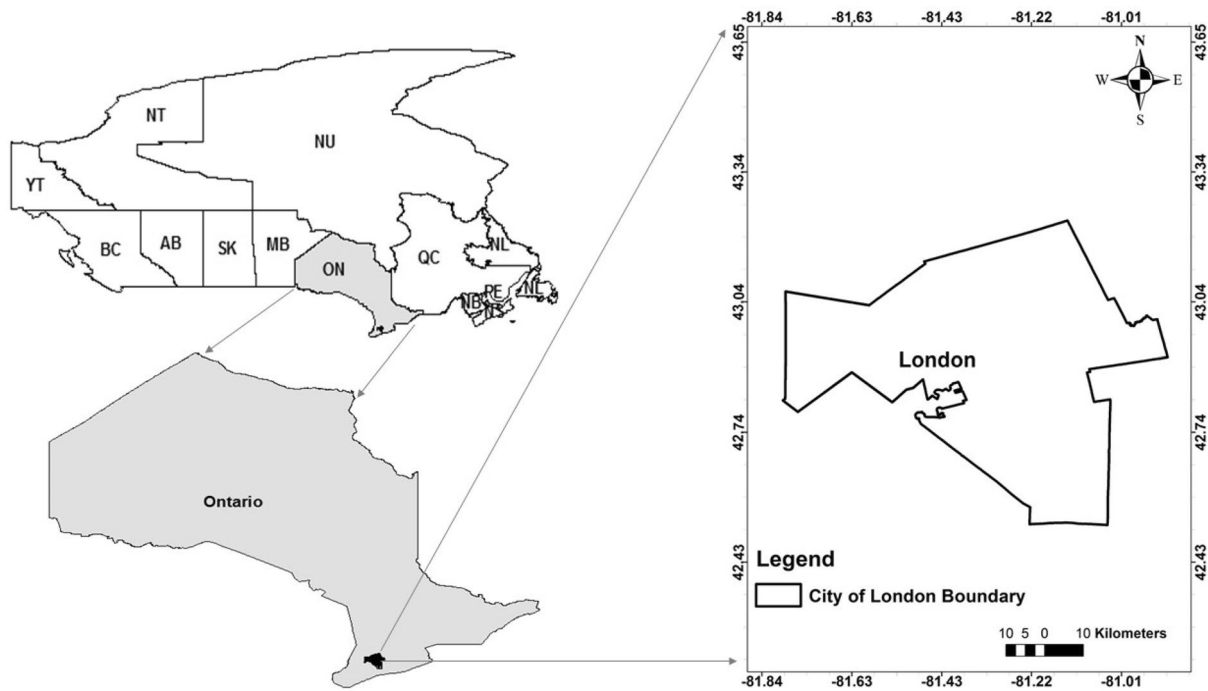**Fig. 1** Time series of the daily water demand for the city of London, Canada

Fig. 2 Study area map showing the city of London, Canada

Table 1 Descriptive statistics of the water use demand data used in the London city of Canada

Metrices	Values
Max (ML/day)	135.78
Min (ML/ day)	85.05
Avg (ML/day)	105.41
SD (ML/day)	9.48
Population	404,699
Time period	2009–2020
Time lag used (days)	1–3–6–12
Data points	4082
Training (%)	75
Testing (%)	12.5
Validation (%)	12.5

output sequences while making a decision. Figure 3 shows an unfolded architecture of simple RNN and the mathematical expressions are as follows:

x_t is the input sequence at a time t. s_t is the hidden memory at the time t, which is estimated based on the current input and the preceding hidden memory. s_t can be formulated as

$$S_t = f(U_{x_t} + WS_{t-1}) \tag{1}$$

where f is the activate function, which acts as a non-linear transformation that results in transferring the input sequence before sending it to the next layer of neurons or concluding it as output. There are different types of activation functions used in deep learning such as sigmoid function, step function, tanh, and rectified linear unit (ReLU). The initial hidden memory s_0 for estimating the first hidden memory s_1 is usually considered 0. o_t is the outcome at a time t and is calculated by

$$o_t = f(VS_t) \tag{2}$$

U, V and W are the weights of the hidden layer, the output layer, and the hidden memory, correspondingly. x_t and o_t are the input and outcome sequences at the time t, respectively.

Despite the advantages of using RNN, it has some disadvantages, such as vanishing gradient and exploding gradient problems [45, 46] and inefficiency of RNN to handle much older past information even using a different activation function.

Long Short-Term Memory (LSTM) as a Form of RNN

LSTM is an upgraded kind of RNN that was designed to model sequential data and their long-term dependencies more precisely than the traditional RNN. The LSTM was designed by Hochreiter, Schmidhuber [47]. It has been so designed that it can overcome the vanishing gradient

Fig. 3 Typical recurrent neural network structure
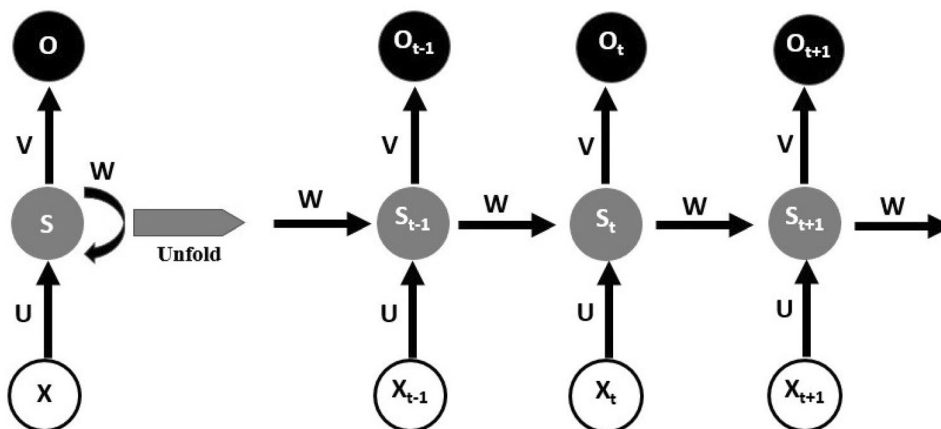
problem. The LSTM architecture consists of a memory cell and three multiplicative gating units. An input gate enhances information to the cell state memory, a forget gate, which eliminates the data that are no longer essential to the model, and an output gate, which picks the information to be shown as output. Figure 4 shows a simple LSTM architecture.

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

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{4}$$

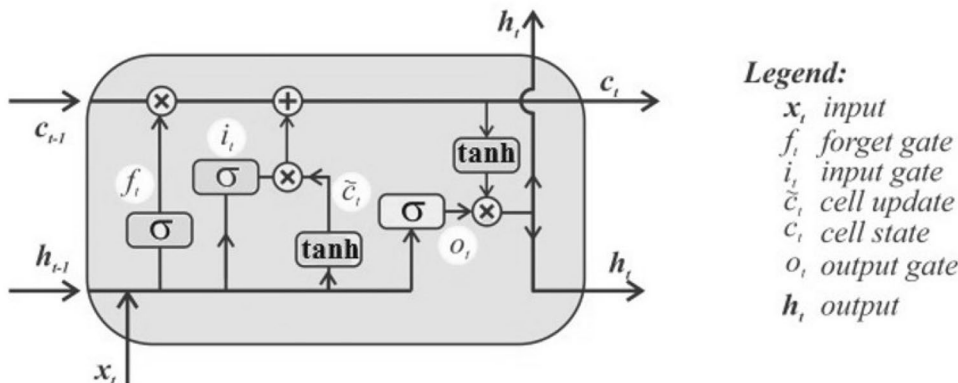$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \tanh(C_t) \tag{8}$$

In the above equations, all weights (W) and biases (b) are learned during the model development period.

One-Dimensional Convolutional Neural Network (1D-CNN)

Convolutional Neural Network (CNN) is a DL neural structure generally used for image recognition [48, 49]. A CNN is made up of a series of convolutional layers, the output of which is only linked to local regions in the input, allowing the network to learn filtering that detects certain patterns in the input data [45, 49]. An architecture of CNN for one-dimensional time series for water demand prediction has been used in this current study (Fig. 5).

A 1D-CNN consists of a convolutional hidden layer that works over a one-dimensional (1D) sequential input layer. Under the situations such as very long input sequences, a second convolution layer is formed and then a pooling layer is generated following the 1D sequential layer whose function is to refine the convolution layer outputs to the most relevant input variables. A dense fully connected layer follows the convolutional and pooling layers which interprets the features extracted by the convolutional layer of the model. A flatten layer is used between the convolutional layers and the dense layer to reduce the feature maps to a single one-dimensional vector. In this study we have developed a convolutional layer with 64 filters, the kernel of size 2, and the Rectifier linear unit (ReLU) as the activation function. The

Fig. 4 LSTM architecture

Legend:
x_t input
f_t forget gate
i_t input gate
$\tilde{c}_t$ cell update
c_t cell state
o_t output gate
h_t output

input feature is then inferred by a max-pooling layer and a dense layer; and finally, the output layer predicts a single numerical value. The model is built-in with the efficient Adam optimization algorithm and the mean-squared error, or 'MSE,' is considered as the loss function.

BiLSTM Model

The conventional LSTM works with the one-directional data processing which has been the factor for reduced efficiency of LSTM model in prediction of water demand [12]. However, multidirectional data may contain valuable information. To overcome these challenges, the bi-directional

LSTM (BiLSTM) incorporates the sequential information from both forward and reverse directions in the dataset, as depicted in Fig. 6.

This means that the past data information of input data sequence is received by forward LSTM whereas future data points are achieved by reverse LSTM [50]. The BiLSTM architecture is incorporated with forget gate structure that is similar to LSTM architecture but the assigned weights and biases in BiLSTM increases two times in comparison with LSTM. In each directions, datasets are trained separately during the training process and finally fused by integrating the outputs as expressed as follows:

$$O_t = \overrightarrow{O_t} \int \overleftarrow{O}_{n-t+1} \tag{9}$$

where $\overrightarrow{O_t}$ and $\overleftarrow{O}_{n-t+1}$ are the output of forward and backward directions, respectively; $\int$. is used for integration operator; and O_t is the predicted output at time t

Hybrid CNN–LSTM and CNN–BiLSTM Models

When advanced ML models such as CNN, LSTM and BiLSTM are used individually, the network structure have their specific advantages. Previous studies suggest that CNN and LSTM (as well as BiLSTM) models could be better suitable stand-alone model for developing a hybrid approach [12]. The long short-term memory (LSTM) model has been evolved to solve the long-term data dependencies due to its gradient vanishing and exploding issues. In this study, hybrid form of DL models named as CNN–LSTM and CNN–BiLSTM have been developed by integrating the advantages of these stand-alone models to predict water demand in different cities of Canada. The CNN–LSTM model consisted of the upper layer with the CNN input layer with several
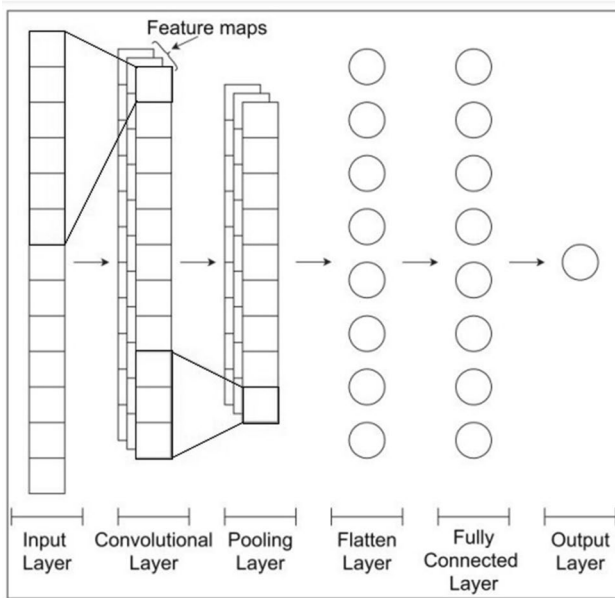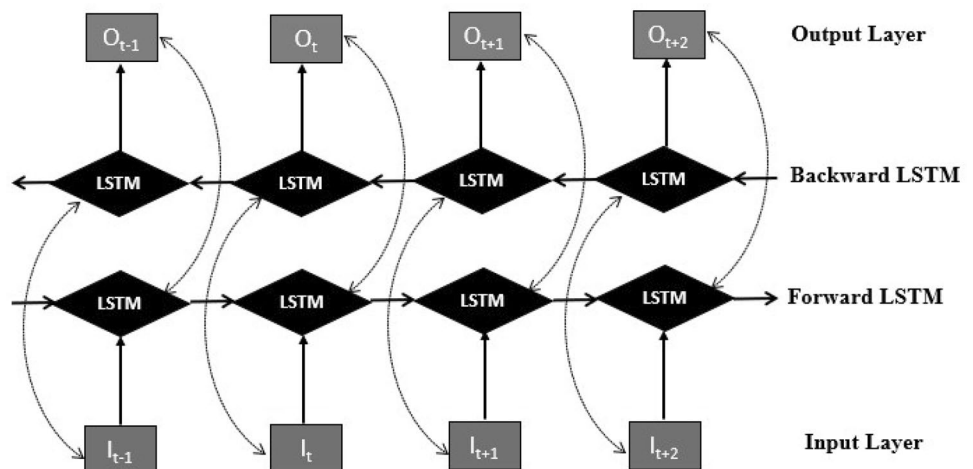

Fig. 5 CNN architecture for water demand prediction model
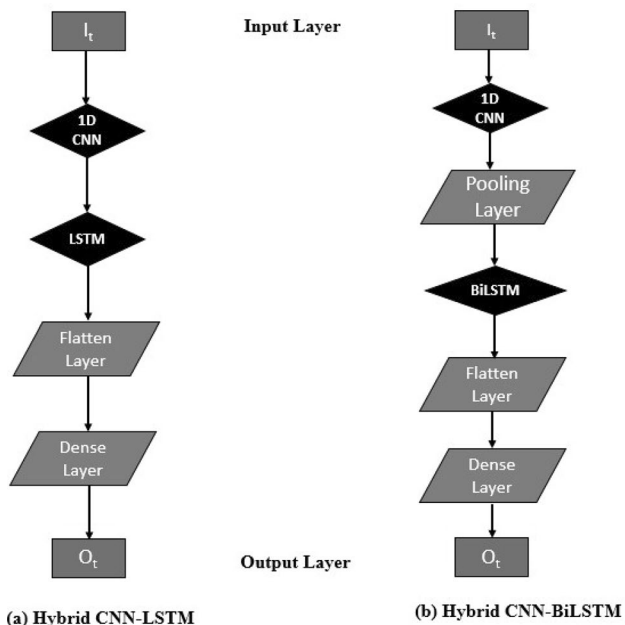
Fig. 6 BiLSTM model architecture

Fig. 7 **a** Hybrid CNN–LSTM and **b** CNN–BiLSTM models

hidden layers and an output layer which extracts features to be given as input to LSTM cells. The hidden layer typically consists of a convolution layer that has already been discussed in the CNN model under "Hybrid CNN–LSTM and CNN–BiLSTM models". The topological architecture of the CNN–LSTM model is presented in Fig. 7a.

The other hybridized form of CNN–BiLSTM model combines the CNN with BiLSTM that merges the benefits from both stand-alone CNN and BiLSTM models. The input vector for CNN model contains the water demand at different time lags and outcome of the CNN model is later fed to the BiLSTM network, and finally, the fully connected layers generate the predicted output. The complete architecture of the proposed CNN–BiLSTM model to forecast water demand in the given framework is shown in Fig. 7b.

Model Development and Hyperparameters Optimization

The stand-alone as well as hybrid models have been developed using Python 3.6.10, Tensorflow 2.1.0 version and Keras deep learning library in Windows 7 [51]. While developing a deep learning model, it is crucial to select a suitable loss function as well as an optimizer. A loss function (or objective function) is one of the essential parameters required to compile a model and to evaluate how well the deep learning procedure fits the observed and predicted variable. There are different types of loss functions such as the cross-entropy loss, mean-squared error (MSE), Huber loss, and hinge loss. However, most of the studies use MSE as a

loss function to train deep learning models used in regression or prediction tasks. We trained the 1D-CNN and LSTM model with Adaptive Moment Estimation (Adam), stochastic gradient descent (SGD), Root Mean Square Propagation (RMSprop), Adaptive Gradient algorithm (AdaGrad), wherein the Adam being an adaptive learning rate optimization procedure performed the best among other optimization algorithms. Tuning the hyperparameters of any deep learning model is a vital part of any model for better input–output mapping. We tuned the hyperparameters of each model based on their Keras packages. Out of the available data from 2009 to 2020, 70% data is used for training, and the rest data are equally divided for validation and testing purposes. Each model has a batch size of 10 and is trained for 500 epochs. As input selection is one of the cumbrous and essential tasks while working with model development, we will follow the input combination through autocorrelation and partial-autocorrelation analysis for the time series. Further, all the DL models used in this study have been developed using the optimal input combination.

Model Performance Evaluation

In the study, the performance of the developed models was evaluated using various statistical measures such as Correlation coefficient (r), Mean Absolute Error (MAE), Nash–Sutcliffe Efficiency (NSE), Scatter Index (SI), Mean Bias Error (MBE), and Discrepancy Ratio (DR) [52–54]:

$$r = \frac{\sum_{i=1}^{N} \left(\mathrm{WD_{Obs}} - \overline{\mathrm{WD_{Obs}}} \right) \left(\mathrm{WD_{Pred}} - \overline{\mathrm{WD_{Pred}}} \right)}{\sqrt{\sum_{i=1}^{N} \left(\mathrm{WD_{Obs}} - \overline{\mathrm{WD_{Obs}}} \right)^2} \sqrt{\sum_{i=1}^{N} \left(\mathrm{WD_{Pred}} - \overline{\mathrm{WD_{Pred}}} \right)^2}} \tag{10}$$

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| \mathrm{WD_{Obs}} - \mathrm{WD_{Pred}} \right| \tag{11}$$

$$\mathrm{MBE} = \left(\frac{1}{n} \right) \sum_{i=1}^{n} \left(\mathrm{WD_{Pred}} - \mathrm{WD_{Obs}} \right) \tag{12}$$

$$\mathrm{NSE} = 1 - \left[\frac{\sum_{i=1}^{N} \left(\mathrm{WD_{Obs}} - \mathrm{WD_{Pred}} \right)^2}{\sum_{i=1}^{N} \left(\mathrm{WD_{Obs}} - \overline{\mathrm{WD_{Obs}}} \right)^2} \right] \tag{13}$$

$$\mathrm{SI} = \sqrt{\frac{\left(\frac{1}{n} \right) \sum_{i=1}^{n} \left(\left(\mathrm{WD_{(Pred)}} - \overline{\mathrm{WD_{(Pred)}}} \right) - \left(\mathrm{WD_{(Obs)}} - \overline{\mathrm{WD_{(Obs)}}} \right) \right)^2}{\left(\frac{1}{n} \right) \sum_{i=1}^{n} \mathrm{WD_{(Obs)}}}} \tag{14}$$

$$\mathrm{DR} = \left[\left(\frac{1}{n} \right) \sum_{i=1}^{n} \frac{\mathrm{WD_{Pred}}}{\mathrm{WD_{Obs}}} \right] - 1 \tag{15}$$

where $WD_{(Obs)}$ is the observed water demand, $WD_{(Pred)}$ is the predicted water demand, $\overline{WD}_{(Obs)}$ and $\overline{WD}_{(Pred)}$ are the mean of the observed and predicted water demand and N is the number of data points.

Uncertainty in Water Demand Utility

The probabilistic forecast would be important in management of water distribution systems involving the confidence level in the advance prediction of water quantity variables. The uncertainty analysis of any model outputs can be made using quantile regression analysis. The predictions of water demand variables in normalized form can be represented in a linear relationship with the residual as follows [55, 56]:

$$NR = m \times ND + c \tag{16}$$

where m and c are the slope and intercept parameters. While doing the regression analysis, the sum of the weighted absolute residuals is minimized using an objective function,

$$\min \sum_{i=1}^{n} \rho_\varepsilon \left| NR_\varepsilon - \left(m_\varepsilon ND_\varepsilon + c_\varepsilon \right) \right| \tag{17}$$

where ε is the desired quantile, and ρ_ε is the quantile regression function for adjustment of NR_ε.

Quantile regression was carried out using the programming functions '*quantilereg*', in the MATLAB package. Using the calibration dataset, the regression between NR and ND at the quantiles of $\varepsilon = 95\%$, 5%, 75% and 25% were analyzed to obtain the regression lines for NR_ε. The m and c parameters were imposed on forecasted discharge to obtain the residuals in the Gaussian domain, NR_ε at different quantiles. The combination of estimated error quantile in the original domain, R_ε and the forecasted WD was obtained as follows:

$$WD_\varepsilon = WD + R_\varepsilon \tag{18}$$

This finds the relationships at desired quantiles for the calibrated data in the original domain; and it can be imposed on any forecasted value by means of linear interpolation or, if forecasted values are found outside the domain of the

Fig. 8 Autocorrelation and Partial autocorrelation analysis for the water demand time series
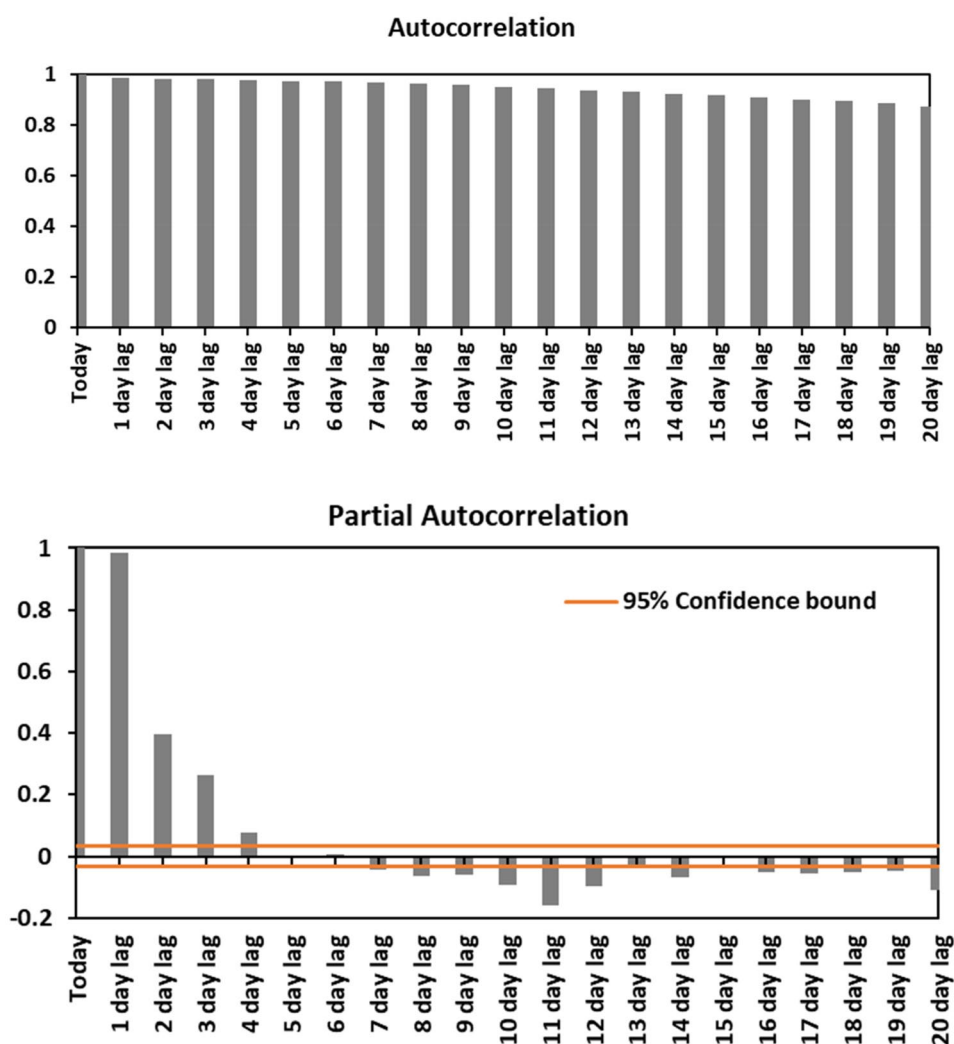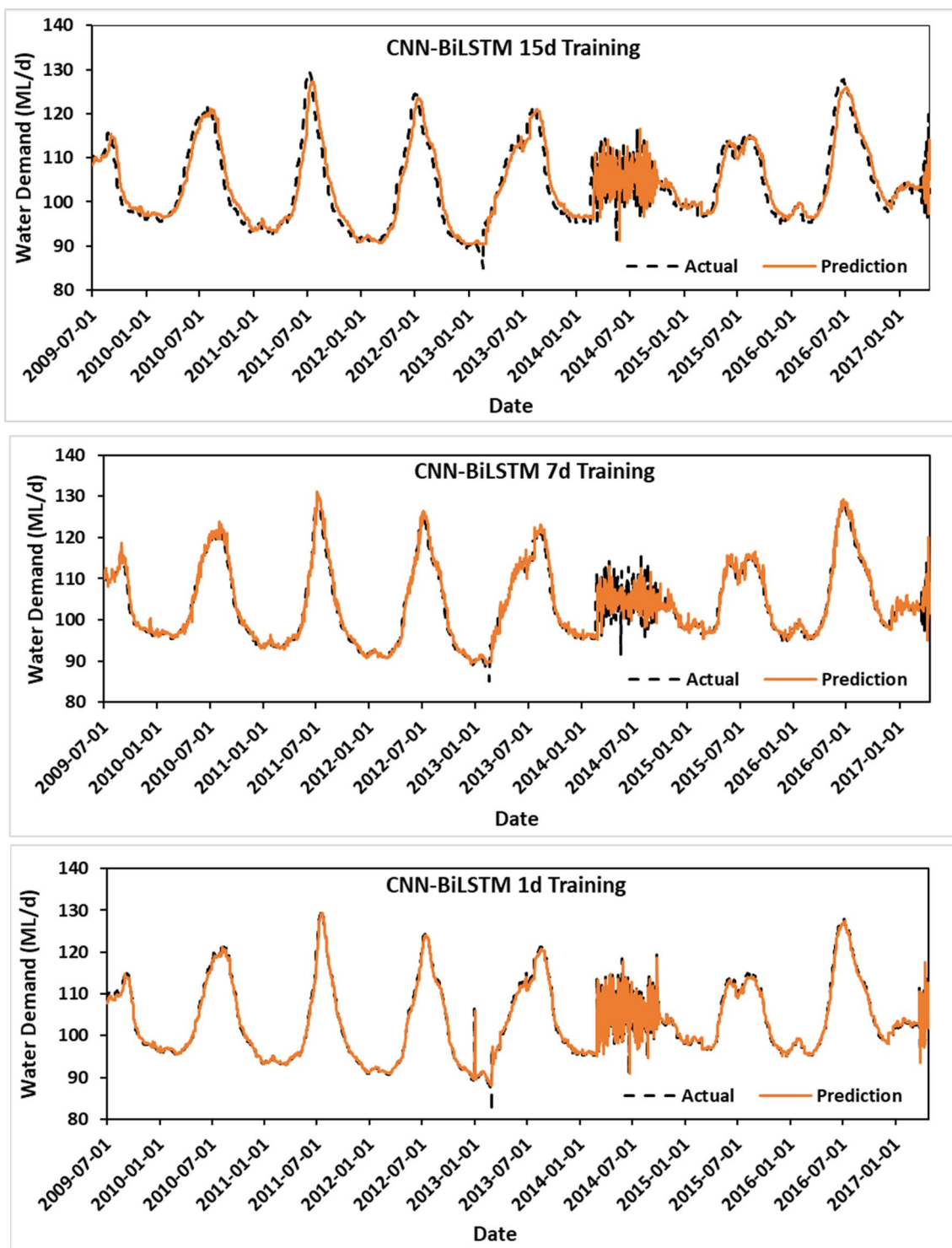
Fig. 9 Training performance of the CNN–BiLSTM model

calibration dataset, with linear extrapolation. Moreover, similar models were derived for several lead times forecasts by the developed models.

Results and Discussion

With the availability of time series of water demand for 12 years, it was first taken towards check of data stationarity.
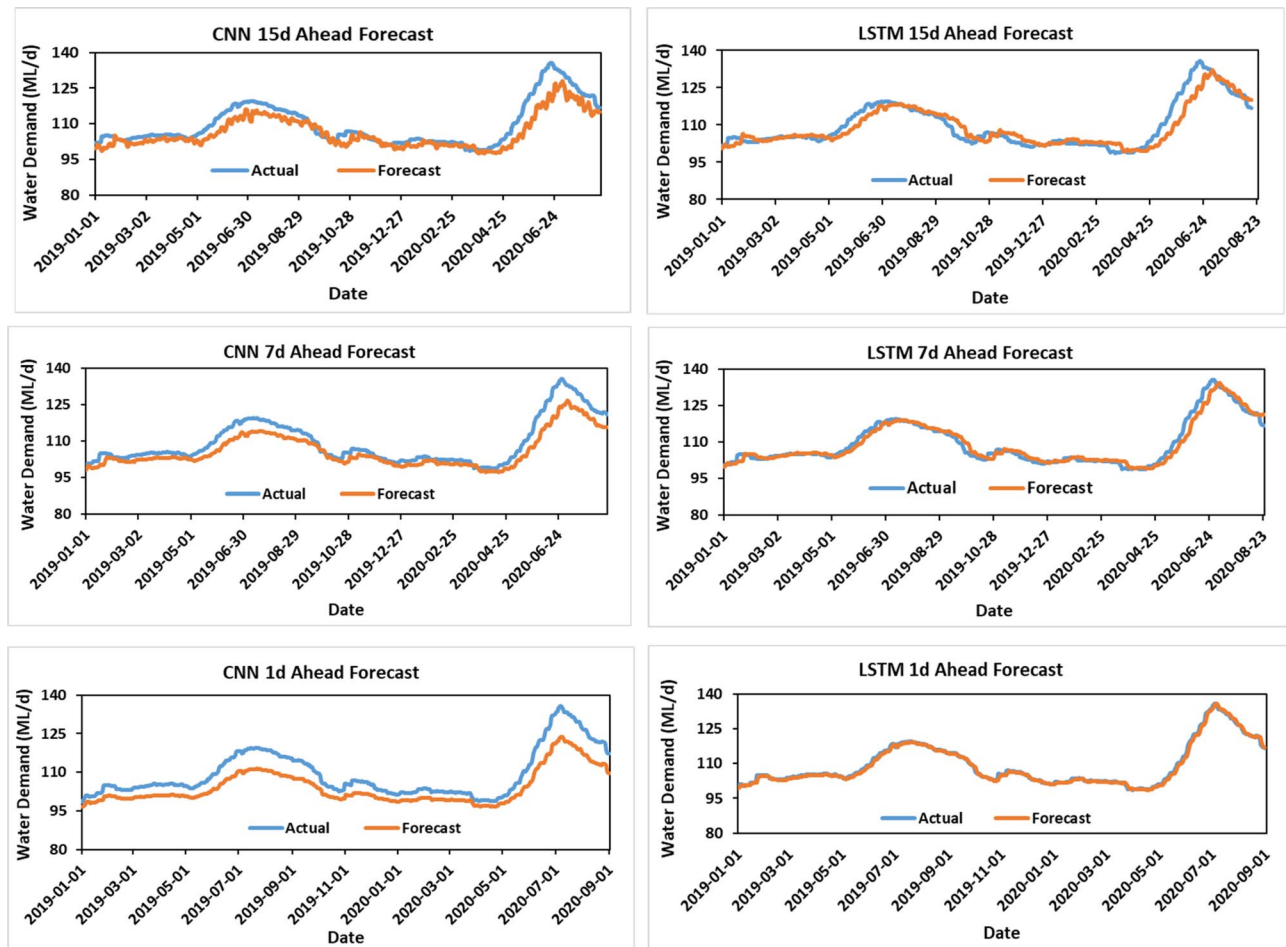
Fig. 10 Visualization of forecasting performance of LSTM over the CNN model

The Augmented Dickey Fuller test, a form of Dickey Fuller test [57], was conducted (ADF Statistic: -5.913) which indicated the p value $< 5\%$ (2.605e$-$07) and the null hypothesis that the data are non-stationary was rejected. Hence, without doing any pre-processing, the existing correlation in the univariate time series of daily water demand was analyzed to find autocorrelation and partial autocorrelation in the time series with respect to time-lag (Fig. 8). The reducing trend of autocorrelation and the partial autocorrelation with 95% confidence bound indicated the dominance of autoregressive (AR) component in the dataset. Hence, the previous day water demand data for 4 days lags were used as the inputs, i.e., from $WD_o (t-1)$ upto $WD_o (t-4)$ into the forecasting models. WD_o is the observed water demand for different time-lag, and WD_f will be the notation for water demand forecast from the model.

Strategic Modeling for Water Demand Forecasting

Initially, the models are trained with observed daily water demand of day 1 $WD_o (t)$ for an output water demand for next day i.e., $WD_f (t+1)$. Further, the models are trained and tested for forecasting at $WD_f (t+1)$, $WD_f (t+7)$ and $WD_f (t+15)$ days which makes the short-term demand forecasts (week advance) essential for management of water network systems.

The model training was conducted with hyperparameter tuning for the CNN, LSTM, BiLSTM and their hybrid models. The time series of model training data and corresponding prediction can be found in Fig. 9 (only CNN–BiLSTM is included). The time series of observed and forecasted water demand during testing phase for both stand-alone CNN and LSTM models are shown in Fig. 10. Figure 11 specifically displays the time series for CNN–BiLSTM models at different lead times throughout the testing period. The forecast accuracy achieved by the developed stand-alone models using training, validation and testing dataset are summarized

Fig. 11 Improved forecasting performance of CNN–BiLSTM model at different lead times
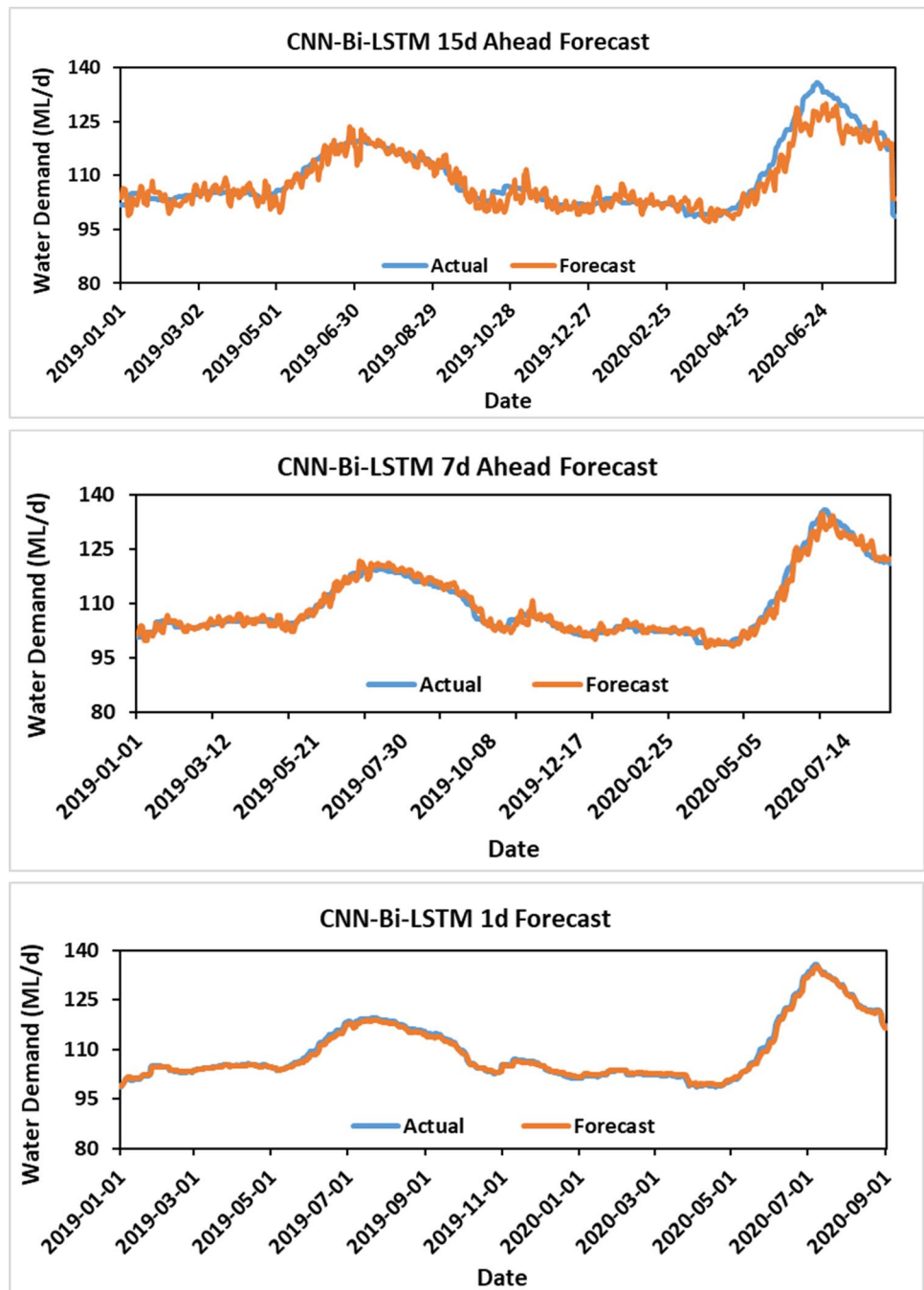
in Table 2. On the other hand, Table 3 provides the forecast accuracy for the hybrid models. Furthermore, performance measures such as SI, MBE, and DR were utilized to accurately assess the prediction of water demand across all the models [54, 58]. The models consistently performed well during training, but in contrast, the CNN model showed efficiency < 90% during 15-day lead. The other models captured the continuous events of daily water demand with NSE > 90% even up to 15-day lead forecasting mode in terms of peak demand and time of the event.

From tables, it is observed that the observed values are closely captured during testing at 1-day-ahead forecasting in case of all the stand-alone DL models. The performance of all stand-alone DL models is as follows: MAE = 0.487 ML/day, NSE = 99.41%, r = 0.998, SI = 0.563, MBE = − 5.288, and DR = 0.049 for CNN model, MAE = 0.537 ML/day, NSE = 99.46%, r = 0.999, SI = 0.064, MBE = − 0.49, and DR = 0.004 for LSTM, and MAE = 0.514 ML/day, NSE = 99.44%, r = 0.999, SI = 0.065, MBE = − 0.469, and DR = 0.004 for the

Table 2 The stand-alone model performance for water demand forecasts

Lead-time of forecast	Models' performance	NSE (%)	r	MAE (ML/day)	SI	MBE	DR
CNN							
1-day	Train	98.21	0.991	0.621	0.482	− 3.694	0.035
	Validation	98.69	0.994	0.562	0.579	− 5.512	0.052
	Test	99.41	0.998	0.487	0.563	− 5.288	0.049
7-day	Train	78.83	0.888	3.006	0.325	− 1.95	0.018
	Validation	57.31	0.757	4.479	0.406	− 3.336	0.03
	Test	65.34	0.808	4.448	0.57	− 6.706	1.057
15-day	Train	68.37	0.827	3.616	0.356	− 1.664	0.015
	Validation	37.36	0.611	5.155	0.433	− 2.979	0.027
	Test	44.47	0.667	5.202	0.444	− 3.242	0.03
LSTM							
1-day	Train	96.09	0.980	0.883	0.211	− 0.2	0.002
	Validation	98.81	0.995	0.570	0.098	− 0.5	0.004
	Test	99.46	0.999	0.537	0.064	− 0.49	0.004
7-day	Train	92.62	0.966	1.602	0.225	0.22	− 0.002
	Validation	91.25	0.958	1.451	0.222	0.05	− 0.001
	Test	95.51	0.979	1.327	0.183	− 0.23	0.002
15-day	Train	81.63	0.920	2.737	0.345	0.37	− 0.004
	Validation	76.03	0.894	2.637	0.355	− 0.09	0.001
	Test	81.65	0.921	2.519	0.329	− 0.63	0.005
BiLSTM							
1-day	Train	96.12	0.980	0.840	0.210	− 0.154	0.001
	Validation	98.79	0.995	0.580	0.099	− 0.748	0.004
	Test	99.44	0.999	0.514	0.065	− 0.469	0.004
7-day	Train	92.58	0.966	1.642	0.241	0.328	− 0.003
	Validation	91.18	0.958	1.487	0.234	− 0.006	0.001
	Test	95.63	0.979	1.336	0.212	− 0.274	0.003
15-day	Train	81.82	0.920	2.736	0.373	0.451	− 0.005
	Validation	76.04	0.894	2.650	0.371	− 0.192	0.002
	Test	81.74	0.921	2.518	0.375	− 0.741	0.006

BiLSTM model. During the testing period, the performance of the hybrid models show MAE = 0.320 ML/day, NSE = 99.734%, r = 0.999, SI = 0.055, MBE = 0.101, and DR = − 0.001 for CNN–LSTM model; and the CNN–BiLSTM model achieved a performance with the following indices: MAE = 0.245 ML/day, NSE = 99.830%, r = 0.999, SI = 0.064, MBE = -0.209, and DR = 0.002. The performance of the models based on NSE varies in the order as follows: CNN–BiLSTM > CNN–LSTM > CNN > BiLSTM > LSTM for training. During validation, the order is CNN–LSTM > LSTM > BiLSTM > CNN–BiLSTM > CNN, and it is CNN–BiLSTM > CNN–LSTM > LSTM > BiLSTM > CNN in testing phase. All the predictive models considered herein are reliable as the NSE values are in the acceptable range of 0.75–1. Moreover, the hybrid model (CNN–LSTM) exhibit improved SI and DR compared to the stand-alone models indicating more precise evaluation of water demand forecast model by the hybrid approaches.

Afterwards, the input scenario with WD (t-1) to WD (t-4), i.e., water demand of 1 day to 4 days before were considered in building and re-training the models. All the selected models were customized for multi-step ahead water demand forecasting and the lead-time water demand as output were simulated using the trained models. In this forecasting task, we have summarized the models for 1-day, 7-day and 15-day lead-time water demand prediction (see Tables 2 and 3). The forecasting performance of the CNN model is not significant due to loss of its data capturing capacity as the model has no memory structure, although it has strong feature selection layers. However, the recurrent-type network of LSTM, BiLSTM as well as the hybrid models with CNN combines the memory structure and performs well in the forecasting problems. For 7-day-ahead water demand forecasts, the NSE of the CNN model reduced to 65.34%, and in case of the other models the NSE values are: for LSTM = 95.51%,

Table 3 The hybrid model performance for water demand forecasts

Lead-time of forecast	Models' performance	NSE (%)	r	MAE (ML/day)	SI	MBE	DR
CNN–LSTM							
1-day	Train	98.892	0.995	0.459	0.026	0.033	−0.786
	Validation	98.893	0.994	0.447	0.091	0.045	−0.001
	Test	99.734	0.999	0.320	0.055	0.101	−0.001
7-day	Train	93.421	0.967	1.556	0.172	−0.112	0.001
	Validation	91.787	0.958	1.483	0.154	−0.541	0.005
	Test	95.641	0.978	1.301	0.154	−0.541	0.005
15-day	Train	84.982	0.922	2.617	0.249	0.313	−0.003
	Validation	79.405	0.891	2.664	0.272	−0.423	0.003
	Test	83.601	0.914	2.483	0.271	−0.422	0.003
CNN–BiLSTM							
1-day	Train	99.290	0.996	0.331	0.034	−0.110	0.001
	Validation	98.740	0.994	0.364	0.099	−0.243	0.002
	Test	99.830	0.999	0.245	0.064	−0.209	0.002
7-day	Train	93.315	0.966	1.634	0.164	0.352	0.003
	Validation	91.545	0.957	1.603	0.229	0.474	−0.004
	Test	95.655	0.978	1.355	0.148	0.112	0.001
15-day	Train	84.799	0.921	2.717	0.365	−0.422	0.004
	Validation	79.586	0.892	2.771	0.298	−0.386	0.004
	Test	84.843	0.921	2.541	0.272	−0.422	0.003

for BiLSTM = 95.63%, for CNN–LSTM = 95.641%, and for CNN–BiLSTM = 95.655%. The stand-alone BiLSTM model is almost equally efficient with the CNN–LSTM hybrid model in case of 7-day- and 15-day-ahead forecast which indicates the robustness of the BiLSTM structure as a stand-alone multi-step forecasting model. The CNN–BiLSTM model produced 15-day forecast with NSE = 84.843% followed by BiLSTM > CNN–LSTM > LSTM > CNN models. The MAE of the forecasts by the CNN–BiLSTM model ranges as 0.245–2.541 ML/day for 1- to 15-day lead which reflects the consistency of the model to long-lead as the model keeps the memory layers updated. The MAE for other models for 1–15 days forecasts range as: CNN–LSTM: 0.320–2.483 ML/day, BiLSTM model: 0.514–2.518 ML/day, LSTM: 0.537–2.519 ML/day, and CNN: 0.487–5.202 ML/day. Similarly, the SI of CNN–LSTM and CNN–BiLSTM models ranges from 0.055 to 0.271 and 0.064 to 0.272, respectively. It can be observed that SI values consistently increase from the 1-day forecast to the 15-day forecast period. This trend is indicative of a higher level of scatter in the data as the forecast horizon extends that can be interpreted from the uncertainty plots (Fig. 12 (b–d)). The results are in-line with the suggestion by Hu et al. [12] as the hybrid approach of CNN and LSTM/BiLSTM models produced satisfactory results in forecasting. In order to assess the uncertainty of the models for lead-time forecasting, quantile regression method has been applied using the training errors to find the parameters (Fig. 12a). The range

of quantiles between 0.95 and 0.05 indicate a 90% CI band; and similarly, the range between 0.75 and 0.25 quantiles indicate 50% CI band.

For representing the resulting uncertainty in streamflow forecasting (for testing results), the forecasts are shown in the form of 90% CI and 50% CI in the area plot; and the corresponding observed water demand time series are shown in red points (Fig. 12b–d). The average width of confidence interval bands is an indicator of the level of uncertainty. It can be observed that the width of CI bands is very narrow at 1-day ahead forecast due to the accurate prediction of the next day water demand forecasting. At higher lead-time modeling by the CNN–BiLSTM model, though the width of the CI band increases, it accommodates most of the observed values within the 90% CI band which indicates the reliability of the model. The CNN–BiLSTM model at 15-day lead outperformed other stand-alone and hybrid CNN–LSTM model. The uncertainty bands of the CNN–BiLSTM show the observed water demand within the CI bands with some underprediction during the peak water demand in year 2020. The number of data points within the bound of the CI varies from 88 to 90% even up to 15-day lead forecasting.

Comparison of the Present Study with Previous Investigations

Comparing our study to previous investigations in the field of urban water demand prediction reveals several novel

Fig. 12 **a** Plot of residuals *vs* forecasts during model training and **b–d** uncertainty plots for the CNN–BiLSTM forecasting
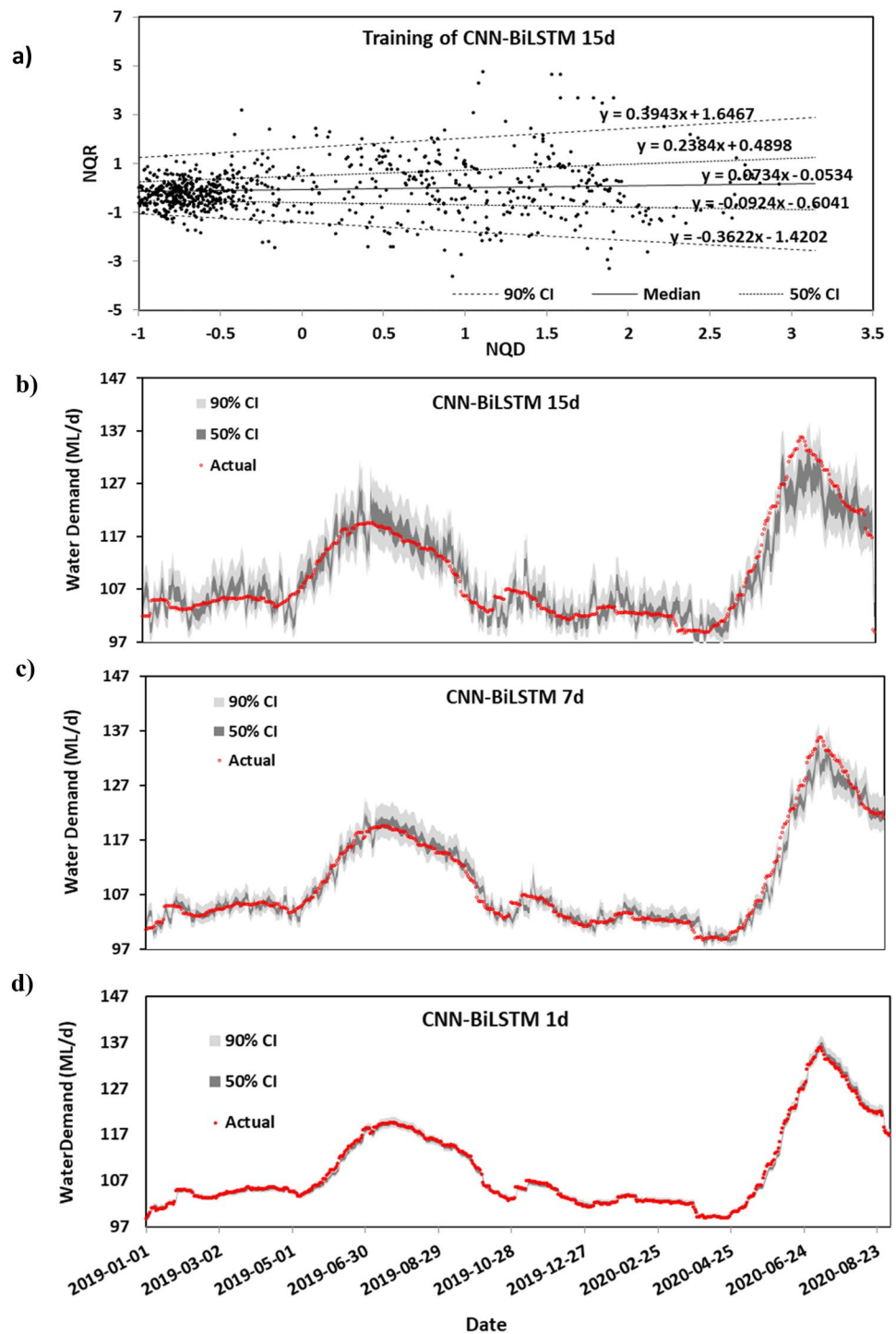
contributions and advancements [14, 59–61]. In contrast to earlier studies that predominantly relied on traditional statistical models or simpler machine learning algorithms, our research employed advanced deep learning models, specifically the CNN–LSTM and CNN–BiLSTM architectures [12]. This choice allowed us to capture complex spatial and temporal dependencies in the water demand data. Moreover,

our study showcased superior performance metrics, including higher *r*, lower MAE, and improved NSE, particularly for shorter lead times [17, 33, 34]. In addition, other metrics such as SI, MBE, and DR were employed to precisely evaluate the prediction of water demand [54, 58]. These results can be attributed to the utilization of extensive and high-resolution datasets, which include additional relevant variables such as

precipitation, temperature, atmospheric pressure, dewpoints, and humidity. Furthermore, these outcomes demonstrate the effectiveness of leveraging the strengths of deep learning techniques [12]. By surpassing previous investigations, our study contributes to the advancement of urban water demand prediction and highlights the potential of deep learning models in achieving accurate and reliable forecasts. Nonetheless, it is important to consider the generalizability of our findings to other regions and time periods.

Conclusions

Urban water demand forecasting is an essential component in the effective management of available water in an urban city. It can help the water managers in decision-making about water demand and supply in a city. In this study, we attempted to forecast one day in advance future water intake in an urban region of the Canadian city, London. We developed a novel hybrid CNN–BiLSTM model for water demand forecasting. Further, the efficacy of the hybrid models is compared with other DL models, viz., LSTM, BiLSTM, and 1D-CNN models. A detailed experimental comparison of these forecasting procedures stepwise at different leads has been investigated using our case study data. Finally, the outcomes indicate that the hybrid CNN–BiLSTM model produces the most accurate forecast, closely followed by the CNN–LSTM, BiLSTM, LSTM, and CNN models. The multi-step forecasting performance of the proposed hybrid CNN–BiLSTM model is due to the powerful feature extraction capability by 1D-CNN model along with Bidirectional LSTM approach that allows for better learning by training two LSTM instead of one LSTM on the input sequence. The model uncertainty aids to the understanding of the model for better acceptability of the point forecasts. Overall, the results support the CNN–BiLSTM approach as a promising deep learning technique for accurate lead-time forecasting of urban water demand.

Data Availability Upon reasonable request, the corresponding author will provide access to all the data and codes that support the findings of this study.

Declarations

Conflict of Interest The authors affirm that they have no conflicts of interest to disclose.

References

1. Nasseri M, Moeini A, Tabesh M. Forecasting monthly urban water demand using extended Kalman filter and genetic programming. Expert Syst Appl. 2011;38(6):7387–95.
2. Walker D, et al. Forecasting domestic water consumption from smart meter readings using statistical methods and artificial neural networks. Procedia Eng. 2015;119:1419–28.
3. Zhou S, et al. Forecasting operational demand for an urban water supply zone. J Hydrol. 2002;259(1–4):189–202.
4. Li W, Huicheng Z. Urban water demand forecasting based on HP filter and fuzzy neural network. J Hydroinf. 2010;12(2):172–84.
5. Ghiassi M, Zimbra DK, Saidane H. Urban water demand forecasting with a dynamic artificial neural network model. J Water Resour Plan Manag. 2008;134(2):138–46.
6. Jain A, Varshney AK, Joshi UC. Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. Water Resour Manag. 2001;15(5):299–321.
7. Adamowski JF. Peak daily water demand forecast modeling using artificial neural networks. J Water Resour Plan Manag. 2008;134(2):119–28.
8. Gato S, Jayasuriya N, Roberts P. Forecasting residential water demand: case study. J Water Resour Plan Manag. 2007;133(4):309–19.
9. Gato S, Jayasuriya N, Roberts P. Temperature and rainfall thresholds for base use urban water demand modelling. J Hydrol. 2007;337(3–4):364–76.
10. Shabani S, et al. Intelligent soft computing models in water demand forecasting. In: Water Stress in Plants. Croatia: InTech; 2016. p. 99–117.
11. Kofinas D, et al. Urban water demand forecasting for the island of Skiathos. Procedia Eng. 2014;89:1023–30.
12. Hu P, et al. A hybrid model based on CNN and Bi-LSTM for urban water demand prediction. In: 2019 IEEE congress on evolutionary computation (CEC). IEEE; 2019.
13. Braun M, et al. 24-hours demand forecasting based on SARIMA and support vector machines. Procedia Eng. 2014;89:926–33.
14. Brentan BM, et al. Hybrid regression model for near real-time urban water demand forecasting. J Comput Appl Math. 2017;309:532–41.
15. Herrera M, et al. Predictive models for forecasting hourly urban water demand. J Hydrol. 2010;387(1–2):141–50.
16. Msiza IS, Nelwamondo FV, Marwala T. Artificial neural networks and support vector machines for water demand time series forecasting. In: 2007 IEEE international conference on systems, man and cybernetics.IEEE; 2007.
17. Mouatadid S, Adamowski J. Using extreme learning machines for short-term urban water demand forecasting. Urban Water J. 2017;14(6):630–8.
18. Tiwari M, Adamowski J, Adamowski K. Water demand forecasting using extreme learning machines. J Water Land Dev. 2016;28:37–52.
19. Qi C, Chang N-B. System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. J Environ Manag. 2011;92(6):1628–41.
20. Tiwari MK, Adamowski JF. Medium-term urban water demand forecasting with limited data using an ensemble wavelet–bootstrap machine-learning approach. J Water Resour Plan Manag. 2015;141(2):04014053.
21. Zubaidi SL, et al. A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. J Hydrol. 2018;561:136–45.
22. Zubaidi SL, et al. A novel methodology for prediction urban water demand by wavelet denoising and adaptive neuro-fuzzy inference system approach. Water. 2020;12(6):1628.

23. Kumar D, et al. Forecasting monthly precipitation using sequential modelling. Hydrol Sci J. 2019;64(6):690–700.
24. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: a survey. Wiley Interdiscip Rev Data Min Knowl Discov. 2018;8(4): e1253.
25. Tang D, Qin B, Liu T. Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdiscip Rev Data Min Knowl Discov. 2015;5(6):292–303.
26. Balaban S. Deep learning and face recognition: the state of the art. In: Biometric and surveillance technology for human and activity identification XII. International Society for Optics and Photonics; 2015
27. Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. In: Proceedings of the British machine vision conference (BMVC). 2015.
28. Young T, et al. Recent trends in deep learning based natural language processing. IEEE Comput Intell Mag. 2018;13(3):55–75.
29. Salman AG, Kanigoro B, Heryadi Y. Weather forecasting using deep learning techniques. In: 2015 International conference on advanced computer science and information systems (ICACSIS). IEEE; 2015.
30. Gamboa JCB. Deep learning for time-series analysis. arXiv preprint arXiv:1701.01887 (2017)
31. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
32. Chen Y, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens. 2016;54(10):6232–51.
33. Guo G, et al. Short-term water demand forecast based on deep learning method. J Water Resour Plan Manag. 2018;144(12):04018076.
34. Salloom T, Kaynak O, He W. A novel deep neural network architecture for real-time water demand forecasting. J Hydrol. 2021;599: 126353.
35. Du B, et al. Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. Expert Syst Appl. 2021;171: 114571.
36. Xu Y, et al. A novel dual-scale deep belief network method for daily urban water demand forecasting. Energies. 2018;11(5):1068.
37. Xu Y, et al. Hourly urban water demand forecasting using the continuous deep belief echo state network. Water. 2019;11(2):351.
38. Kim J, et al. Development of a deep learning-based prediction model for water consumption at the household level. Water. 2022;14(9):1512.
39. Kavya M, et al. Short term water demand forecast modelling using artificial intelligence for smart water management. Sustain Cities Soc. 2023;95: 104610.
40. Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos Solitons Fract. 2020;140: 110212.
41. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5(4):115–33.
42. Rumelhart D, Hinton G, Williams R. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. Parallel distributed processing. Cambridge: MIT Press; 1986.
43. Caudill M, Butler C. Understanding neural networks; computer explorations. Cambridge: MIT Press; 1992.
44. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.
45. Pascanu R, Mikolov T, Bengio Y (2013) On the difficulty of training recurrent neural networks. In: International conference on machine learning. PMLR; 2013.
46. Kag A, Zhang Z, Saligrama V. RNNS incrementally evolving on an equilibrium manifold: a panacea for vanishing and exploding gradients? In: International Conference on Learning representations. 2019.
47. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
48. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097–105.
49. Chaerun Nisa E, Kuan Y-D. Comparative assessment to predict and forecast water-cooled chiller power consumption using machine learning and deep learning algorithms. Sustainability. 2021;13(2):744.
50. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 2005;18(5–6):602–10.
51. Chollet F. Keras: Deep learning library for theano and tensorflow, vol. 7, no. 8, p. T1. 2015. https://keras.io/k.
52. Shamshirband S, et al. Predicting standardized streamflow index for hydrological drought using machine learning models. Eng Appl Comput Fluid Mech. 2020;14(1):339–50.
53. Sattari MT, et al. Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall. Eng Appl Comput Fluid Mech. 2020;14(1):1078–94.
54. Saberi-Movahed F, Najafzadeh M, Mehrpooya A. Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions. Water Resour Manag. 2020;34(2):529–61.
55. Jeong DI, Kim YO. Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrol Process Int J. 2005;19(19):3819–35.
56. Weerts A, Winsemius H, Verkade J. Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). Hydrol Earth Syst Sci. 2011;15(1):255–65.
57. Fuller WA. Introduction to statistical time series. New York: Wiley; 1976.
58. Amiri-Ardakani Y, Najafzadeh M. Pipe break rate assessment while considering physical and operational factors: a methodology based on global positioning system and data-driven techniques. Water Resour Manag. 2021;35(11):3703–20.
59. Tiwari MK, Adamowski J. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour Res. 2013;49(10):6486–507.
60. Yong L, Jiahong L. A distributed model for urban water demand prediction. Chin Sci Bull. 2017;62(24):2770–9.
61. Huayi L, et al. Research and application of urban water demand forecasting based on time difference coefficient. J Shanghai Jiaotong Univ (Chin Ed). 2017;51(10):1260.