




Deep learning–based short-term water demand forecasting in urban areas: A hybrid multichannel model

Hossein Namdari ^{*}, Seyed Mohammad Ashrafi  and Ali Haghghi 

Faculty of Civil Engineering and Architecture, Shahid Chamran University of Ahvaz, Golestan Blvd, Ahvaz, Islamic Republic of Iran

^{*}Corresponding author. E-mail: h-namdari@stu.scu.ac.ir

 HN, 0000-0003-4454-9145; SMA, 0000-0001-7884-9029; AH, 0000-0002-2765-6929

ABSTRACT

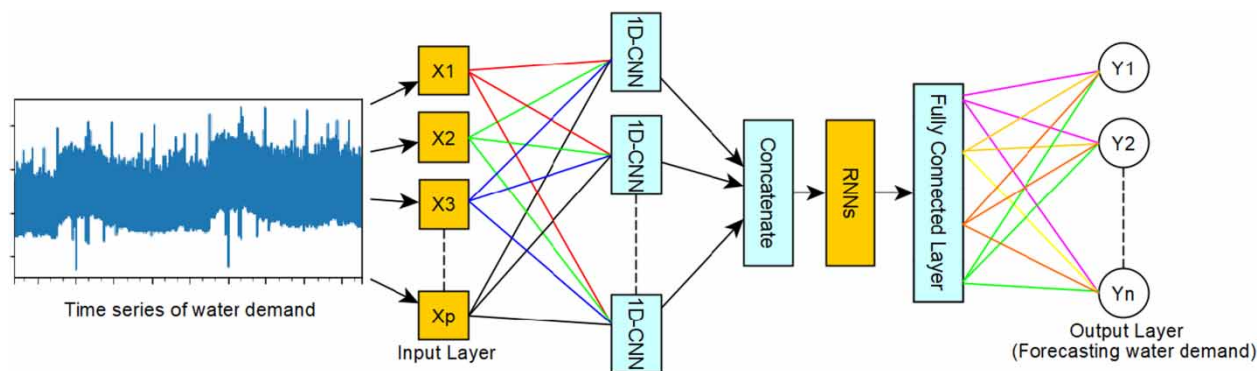
Forecasting short-term water demands is one of the most critical needs of operating companies of urban water distribution networks. Water demands have a time series nature, and various factors affect their variations and patterns, which make it difficult to forecast. In this study, we first implemented a hybrid model of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to forecast urban water demand. These models include a combination of CNN with simple RNN (CNN-Simple RNN), CNN with the gate recurrent unit (CNN-GRU), and CNN with the long short-term memory (CNN-LSTM). Then, we increased the number of CNN channels to achieve higher accuracy. The accuracy of the models increased with the number of CNN channels up to four. The evaluation metrics show that the CNN-GRU model is superior to other models. Ultimately, the four-channel CNN-GRU model demonstrated the highest accuracy, achieving a mean absolute percentage error (MAPE) of 1.65% for a 24-h forecasting horizon. The effects of the forecast horizon on the accuracy of the results were also investigated. The results show that the MAPE for a 1-h forecast horizon is 1.06% in four-channel CNN-GRU, and its value decreases with the amount of the forecast horizon.

Key words: convolutional neural networks, forecasting short-term water demand, GRU, hybrid multichannel deep learning, LSTM, recurrent neural networks

HIGHLIGHTS

- Time series analysis of water demand using hybrid deep learning models can be a suitable option for short-term forecasting.
- Hybrid deep neural networks integrate the advantages of the two classic models of convolutional neural networks (CNNs) and recurrent neural networks (RNNs).
- The combination of CNN and RNNs can simultaneously extract the appropriate features by CNN and learn the long dependency between data by RNNs.

GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

1. INTRODUCTION

Urban water distribution networks (WDNs) are one of the vital infrastructures of any city. Intelligent operation management is necessary for WDNs to ensure adequate water supply at the desired pressure and quantity for consumers. One of the essential requirements of operation management in water supply networks is to predict the amount of water demand needed in that network in the short-term and the hourly interval. Many factors affect the quantity of water demand, such as temperature, precipitation, relative humidity, population, network water pressure, water price for various uses, water losses, the method and system of measuring water consumption, household income, yard area, and green space (Arbués *et al.* 2003; Wentz & Gober 2007; Schleich & Hillenbrand 2009; Nauges & Whittington 2010; de Maria André & Carvalho 2014; de Souza Groppo *et al.* 2019).

Water demand forecasting methods can be divided into two general categories of linear and nonlinear methods (Zhang 2001; de Souza Groppo *et al.* 2019). Exponential smoothing, autoregressive integrated moving averages (ARIMA), and multivariate linear regression are linear methods that use univariate or multivariate time series analysis (Adamowski & Karapataki 2010; Adamowski *et al.* 2012; Ristow *et al.* 2021). In this regard, various regression methods were used by Shuang & Zhao (2021) to predict urban water demand. Short-term water demand follows a nonlinear pattern and is affected by many factors. In most of the past studies, statistical methods were less accurate than nonlinear methods for forecasting short-term water demand. These methods are used more for long-term forecasting (Donkor *et al.* 2014; Ghalekhondabi *et al.* 2017).

Nonlinear regression and artificial neural networks (ANNs) with nonlinear activation functions are nonlinear methods. For example, ANNs were used by Ghiassi *et al.* (2008) to predict water demand. Herrera *et al.* (2010), Peña-Guzmán *et al.* (2016), and Brentan *et al.* (2017) used the support vector machine method. Also, Altunkaynak *et al.* (2005) and Firat *et al.* (2009a) used fuzzy logic. Other ANNs used in past studies to predict urban water demand can be cited, such as the generalized regression neural network, the radial basis function networks, the feedforward neural network (Firat *et al.* 2009b), and the extreme learning machine method (Mouatadid & Adamowski 2017). A machine learning (ML) model such as the abovementioned models cannot simultaneously perform feature selection and prediction. On the other hand, using all the features in ML models, in addition to the problems of collecting data in the real world, increases the computational cost and reduces the model's accuracy. Therefore, to feature selection, reduce the computational cost, adjust neural network parameters, and eliminate noise from data, the researchers focused on hybrid methods for predicting water demand (Tiwari & Adamowski 2015; Shirkoohi *et al.* 2021). As mentioned earlier, one of the significant challenges of using ML methods is choosing the appropriate features that directly affect the prediction results. On top of this, these models often pose an overfitting problem with the increase in data (Sajjad *et al.* 2020). In deep learning methods, feature selections are made automatically through many hidden layers; their accuracy usually increases with increasing data. Similarly, several deep neural networks (DNNs) are developed for water demand prediction. For example, Mu *et al.* (2020) used the long short-term memory (LSTM) model to predict hourly and daily water demand in Hefei in China. They obtained better results than support vector regression, random forest, and ARIMA models. In another study, Guo *et al.* (2018) used the gate recurrent units (GRU) model to forecast water demand with a time step of 15 min. In this study, the GRU model was more accurate than the conventional ANN model and the seasonal ARIMA model. Convolutional neural networks (CNNs) are widely used in various fields, including time series, due to their high ability to extract features, and recurrent neural networks (RNNs) can learn time dependency between data. Namdari *et al.* (2023) used a one-dimensional CNN (1D CNN) to forecast short-term urban water demand. Next, they compared the results with other deep learning models including simple RNNs, LSTM, GRU, and deep feedforward neural networks (DFNNs). The results of their study showed that 1D CNN, compared to other models, predicts short-term water demand with higher accuracy, and the DFNN model has the least accuracy compared to other models.

Hybrid DNNs can consist of CNNs and RNNs. In hybrid DNNs, CNNs are utilized to capture spatial features, while recurrent models are employed to model temporal features (Sajjad *et al.* 2020). The convolution layers in these models extract the features from the dataset, and the RNNs learn the long dependency between the data. Hybrid DNNs excel in recognizing patterns with both spatial and temporal characteristics. Recently, researchers used these models for prediction in various fields, for example, human activity recognition (Lu *et al.* 2022), petroleum price prediction (Kim & Jang 2023), electric energy forecasting (Wang *et al.* 2023b; Yang *et al.* 2023), estimation of the yield of agricultural products (Wang *et al.* 2023a), wind speed forecasting (Lv *et al.* 2023), and forecasting stock market indices (Song & Choi 2023).

In past studies, researchers have used various methods to predict water demand, but despite much research in this field, the hybrid DNN has received less attention. The water demand follows a time series pattern with a complex and multifaceted structure, and it is challenging to collect all the characteristics affecting water demand in a real-world implementation. On the other hand, despite having a high ability to analyze nonlinear water demand data, ML models are weak in dealing with nonstationary data (Ghalehkhondabi *et al.* 2017). In this study, we implemented hybrid DNNs to forecast short-term urban water demand by combining CNN and RNNs. The combination of CNN and RNNs can simultaneously extract the appropriate features by CNN and learn the time dependency between data by RNNs.

2. METHODOLOGY

The pattern of water demand has the nature of a time series. In this study, we seek to find the pattern of water demand based on historical records of water demand to forecast the hourly water demand for the next 24 h. In other words, we want to solve a multistep univariate time series problem using hybrid DNNs. For this purpose, first, the water demand data are preprocessed. The missing values are filled with appropriate values, and outliers are identified and removed from the dataset. The models used in this study are supervised models that need labeled data. To convert the data into labeled data, we obtain the optimal lag between the data by calculating the autocorrelation of the dataset. Then, we convert the data into labeled data based on the optimal data lag, as described in Section 5.2. We combined the 1D CNN with RNNs and implemented it after the dataset preparation. These networks include the following: (1D CNN + Simple RNN), (1D CNN + LSTM), and (1D CNN + GRU). To increase the accuracy of the model, we increased the number of CNN routes (channels). The input data are entered into multiple 1D CNN channels with different filter sizes. The outputs from the CNN channels are concatenated and then passed into RNN layers. Finally, we compared the results of these hybrid models and introduced the best hybrid DNNs to predict short-term water demand. Figure 1 shows the proposed framework of the hybrid DNNs in which the RNN blocks in the models are Simple RNN, LSTM, and GRU. In this figure, $[X_1, X_2, \dots, X_p]$ is the input layer, and $[Y_1, Y_2, \dots, Y_n]$ are the forecast values for the first hour to the last hour of the time horizon. In the following, after the introduction of DNNs, the implementation method will be explained.

3. INTRODUCTION OF DNNs

3.1. 1D CNNs

Two-dimensional CNNs (2D CNNs) are generally used in image and video recognition, image classification, medical image analysis, and natural language processing. These networks are a component of feedforward ANNs that include convolution and pooling layers. A modified version of 2D CNNs called 1D CNNs has recently been developed. The primary difference between these two networks is that 1D arrays are used instead of 2D matrices in convolution and pooling layers to prepare the feature map. The computational cost of 1D CNNs is much lower than that of 2D CNNs (Kiranyaz *et al.* 2021; Qazi *et al.* 2022). In this sense, they are an excellent choice for signal processing and analyzing time series data. Figure 2 shows the details of a 1D CNN.

3.2. Simple RNNs

RNNs belong to the family of DNNs and are specifically designed for processing sequential data. These neural networks have memory created by a recurrent unit in these networks. In other words, these networks, in addition to the standard layers in multilayer neural networks, have a return connection in the hidden layer, which makes it transfer the information of the previous step and play a role in decision-making. According to Figure 3, RNN units can be defined with a recursive relation:

$$s_t = f(x_t, s_{t-1}) \quad (1)$$

In this equation, x_t is the input of the network at time t , s_t is a vector of values called the state of the internal network at time t , and s_{t-1} refers to a summary of previous network inputs for times before t .

3.3. LSTM neural network

LSTMs are a particular type of RNN. Hochreiter & Schmidhuber (1997) created a neural network in 1997 to learn long-term dependencies. This network can learn long-term dependencies in the dataset due to having a special memory cell. There are three types of gates in LSTM: forgetting, input, and output. In the forgetting gate, what information to delete or retain in the

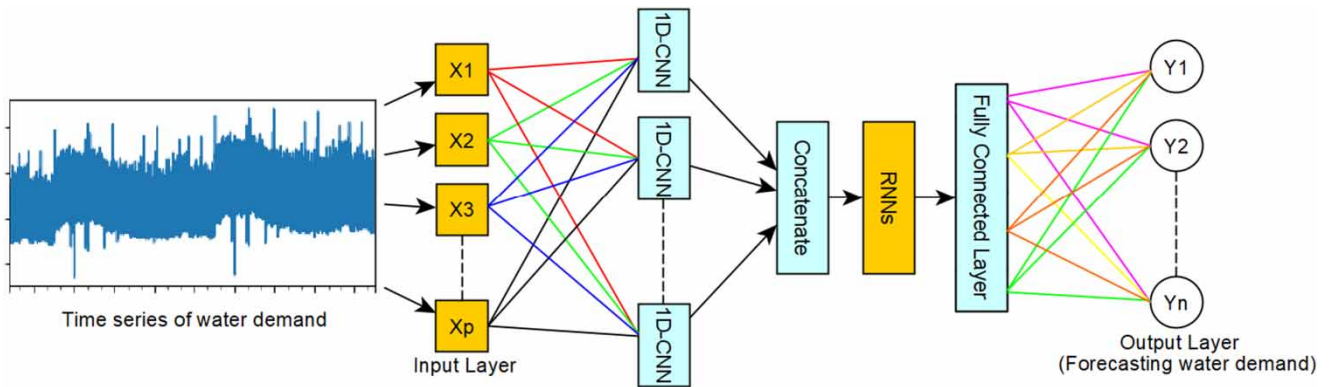


Figure 1 | The proposed framework of the multichannel hybrid DNNs.

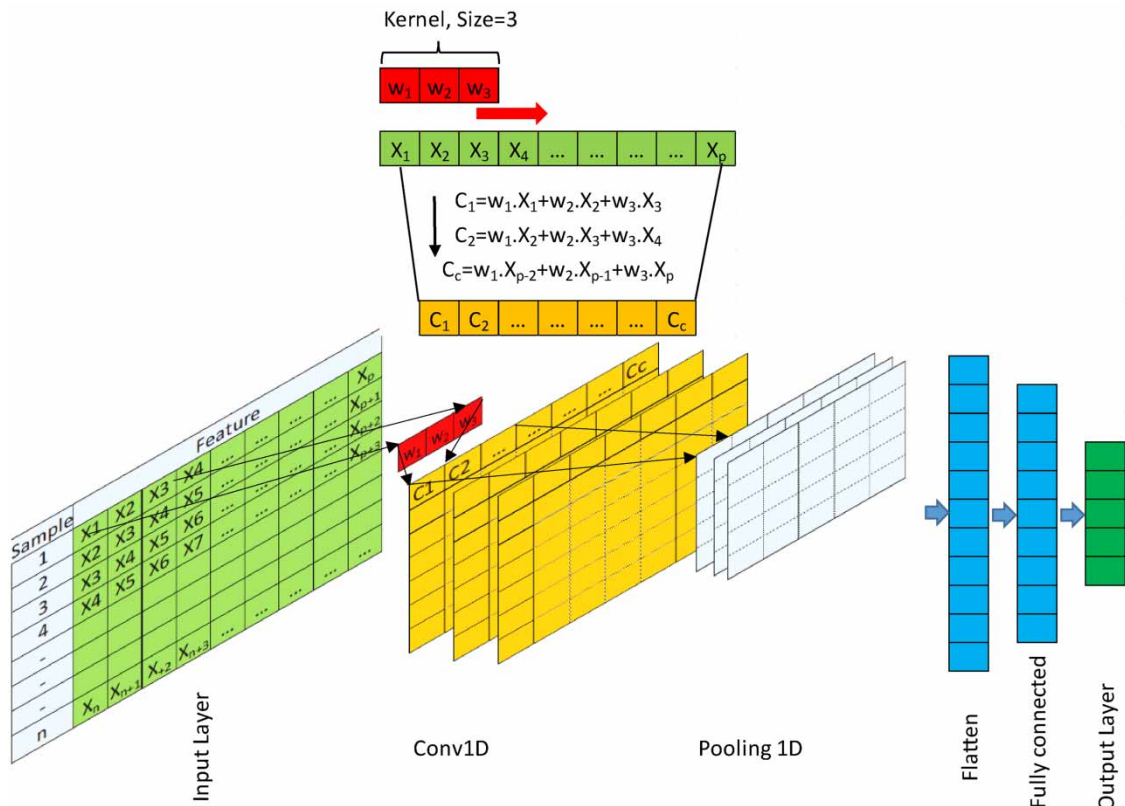


Figure 2 | 1D CNN for time series data analysis.

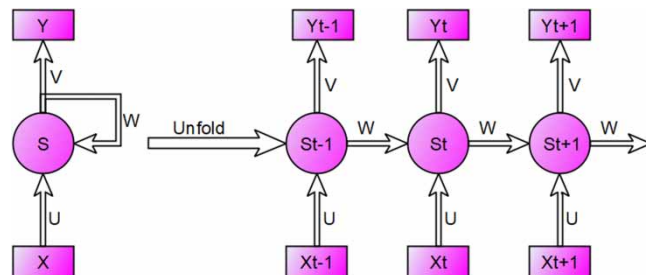


Figure 3 | A simple RNN.

memory cell is decided; in the input gate, the decision is made in the direction of what new information is stored in the state cell; and in the output gate, how much of the long-term memory should be transferred to the output is determined. Figure 4 shows the architecture of the LSTM unit.

3.4. GRU neural network

The GRU architecture was designed by *Cho et al. (2014)*. Two gates are used in the GRU network: update and reset gates. In the update gate, based on the new input and the information in the memory cell related to the past, it decides what information should be kept and what information should be deleted. In the reset gate, based on the new input (at time t) and the information in the memory cell, it decides what information should be deleted. The information from the previous time step's memory cell is combined with the remaining information to generate a candidate memory cell. In the last step, the update gate determines the proportion of the candidate memory information and memory cell information from the previous step that will be added to obtain the memory cell vector at time t . The memory mentioned is utilized as a hidden state for the time $t + 1$, and by applying an activation function on it, the output at time t can be obtained. Figure 5 shows the architecture of the GRU unit.

3.5. Hybrid DNNs

Hybrid DNNs leverage water demand data sequences as input to a convolutional layer structure, which performs feature extraction. The output of the convolutional layer is then passed through a flattening layer and fed as input to RNNs, whose final output is the forecast value of the water demand. Following the experimental comparison, it has been observed that this solution effectively combines the benefits of the two traditional models of CNN and RNNs (*Xu et al. 2021*). The

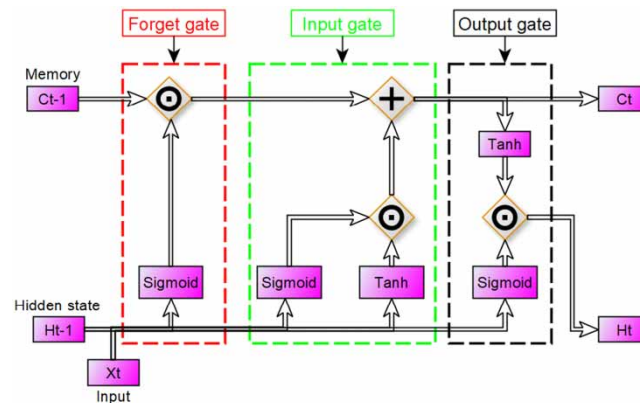


Figure 4 | Architecture of the LSTM unit.

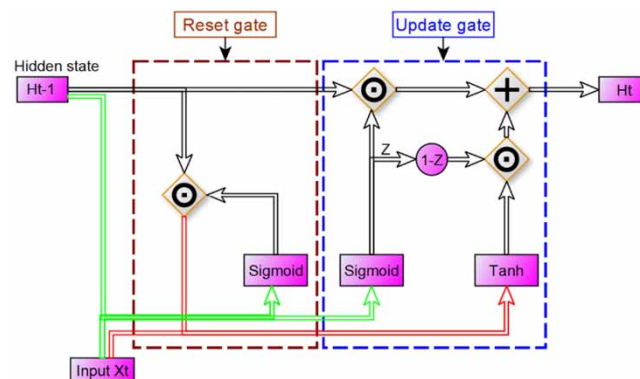


Figure 5 | Architecture of the GRU unit.

CNN layer aids in comprehending the sequential features of the input while the RNN layer learns long-term dependencies between data (Xu *et al.* 2021). Various types of RNNs are used in these models, including Simple RNN, LSTM, and GRU.

4. CASE STUDY

Shiraz City is the capital of Fars province, located in the south of Iran. This city is the fourth most populated in Iran, with a population of about two million people and an area of 1,268 km². This city, based on meteorological data, has a moderate climate. Its average annual temperature is 19.3 °C, and its annual rainfall is 294 mm. According to the reports of Shiraz Water and Wastewater Company, about 75% of drinking water in this city is supplied by 180 underground water wells, and 25% is supplied by surface water. Due to the large number of water production sources and pumping stations in the WDN, it is necessary for short-term water demand forecasts in hourly intervals for operating companies. With a proper forecast of short-term water demand, the operating company can deliver water to the customers in sufficient quantity and at the appropriate pressure and minimize energy consumption costs with proper planning for pump operations. Saadi Town is a mountainous region in the east of Shiraz. It covers an area of 360 hectares and has a population of over 60,000 people. We considered this section of Shiraz City as the study area with an independent WDN. The dataset used in this study was 34,840 hourly water demand records from 19 May 2016 to 18 May 2022, of which 8,984 data records are missing. The minimum and maximum hourly demand in the dataset is 131.7 and 1,103.2 m³ with an average of 596.49 and a standard deviation of 131.7 m³.

5. IMPLEMENTATION OF HYBRID DNNS

5.1. Data preprocessing

Data preprocessing is an important step in data mining projects, and quality decisions can only be made when they are based on quality data. In this research, we used hourly water demand in past years. Water demand data may not be recorded in the dataset at some hours of the day for various reasons. To fill in the missing values, if the missing data are related to an hour on a working day, we used the average demand in the same hour on the day before and after it. In another case, the missing data are related to a holiday or the day before or after a holiday; we used the average demand at the same time and day in the week before and after that. This method for filling missing values is the most likely state for missing values and prevents data bias. Using this method to fill in the missing values due to the high probability of occurrence and being close to reality will considerably solve the problem of bias in the data and also preserve the seasonal changes in the value of demand in correcting the missing value.

Outlier data are data that are significantly different from other data. Their use in ML models causes anomalies and negative bias in the results. We identified outliers using the boxplot. Through this method, we calculate the value of the data's first and third quartiles (Q_1 , Q_3) and get the difference between them ($IQR = Q_3 - Q_1$). Data that are larger than $Q_3 + 1.5IQR$ or smaller than $Q_1 - 1.5IQR$ are considered outliers. In the dataset, 52 hourly records were identified as outliers using the boxplot and removed from the dataset. Then, the deleted values were filled in as in the case of missing data.

Normalization is rescaling data from the original scale so that all values are taken to a small and specific domain. Neural network models use the gradient descent algorithm to minimize the cost function, and if normalization is used, the convergence speed is much higher (Han *et al.* 2022). We used Min–Max normalization in this study and converted the dataset to the range of 0–1 according to Equation (2):

$$V' = \frac{V - \text{Min}_A}{\text{Max}_A - \text{Min}_A} \quad (2)$$

The Min_A and Max_A are the minimum and maximum values of a feature A , V is the value of the feature, and V' is the value of the normalized feature.

5.2. Implementation of single-channel hybrid models

To pump scheduling in WDNs, it is necessary to forecast the hourly water demand for the next 24 h. The pattern of water demand has the nature of a time series. To forecast water demand in the next 24 h, we solve a time series problem using deep supervised learning models. The data need to be labeled for supervised learning models, so we calculate the optimal lag between water demand data using the autocorrelation function. This helps us convert the dataset into labeled data.

We use the following equation to calculate the optimal lag:

$$AC(X_t, X_{t-lag}) = \frac{Cov(X_t, X_{t-lag})}{Var(X_t)} \tag{3}$$

In this equation, $AC(X_t, X_{t-lag})$ represents the autocorrelation, $Cov(X_t, X_{t-lag})$ is the covariance between values at time t and t -lag, and $Var(X_t)$ denotes the variance of values at time t . We identify the optimal lag by drawing the autocorrelation plot of the water demand data. The autocorrelation plot between the data is shown in Figure 6. The blue shaded area in this figure indicates a 95% significant level, and the vertical lines (correlation values) that are taller than this area indicate the presence of significant correlation in that lag (Lazzeri 2020). As evident in this figure, the highest autocorrelation between the data in lag is 24. Hence, we select the water demand data from time 1–24 as the feature of the first sample and its value from time 25–48 as the sample’s label. We repeat this process until the final data (Table 1) are labeled and used as model input (Namdari et al. 2023). Then, we selected 80% of the beginning of the dataset for the training data and 20% of the end of the dataset for the test data.

Hyperparameter tuning was done manually so that we implemented a 1D CNN model with 16, 32, 64, and 128 filters, kernel size = 2–4, and 1–4 1D-convolutional layers. Similarly, for hybrid DNNs, we implemented various combinations of the number of layers and blocks of RNN after 1D CNN layers. By changing and adjusting other hyperparameters (i.e., activation functions, initial weighting method of parameters, number of epochs, and optimization algorithm), we obtained the performance of the models with each combination. We selected the best architecture for each model. The selected 1D CNN model had three layers of 1D convolution (Conv1D) with 64 filters in the first layer, 128 filters in the second layer, 64 filters in the third layer, with kernel size = 3, and a max-pooling layer with size 2. The Rectified Linear Unit (ReLU) activation function was used in all layers. In selected hybrid DNNs, after convolution layers and max-pooling 1D, two layers with RNN blocks were used (Simple RNN, LSTM, and GRU). Then, its output enters a fully connected layer with 24 neurons with a scaled exponential linear unit (SELU) activation function, the forecast values of water demand in the next 24 h. RNN layers

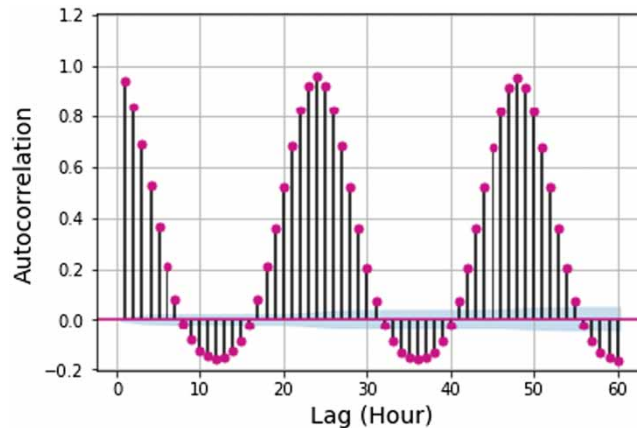


Figure 6 | Autocorrelation plot of hourly water demand data.

Table 1 | Converting water demand time series into the labeled dataset

Sample	Feature					Label				
1	Q_1	Q_2	Q_3	...	Q_{24}	Q_{25}	Q_{26}	Q_{27}	...	Q_{48}
2	Q_2	Q_3	Q_4	...	Q_{25}	Q_{26}	Q_{27}	Q_{28}	...	Q_{49}
3	Q_3	Q_4	Q_5	...	Q_{26}	Q_{27}	Q_{28}	Q_{29}	...	Q_{50}
4	Q_4	Q_5	Q_6	...	Q_{27}	Q_{28}	Q_{29}	Q_{30}	...	Q_{51}
...
n	Q_n	Q_{n+1}	Q_{n+2}	...	Q_{n+23}	Q_{n+24}	Q_{n+25}	Q_{n+26}	...	Q_{n+47}

are in the CNN-GRU model with 100 and 80 GRU blocks with ReLU and SELU activation functions, respectively. The CNN-LSTM model used 50 and 32 LSTM blocks with ReLU activation functions, and the CNN-Simple RNN model used 100 and 50 blocks of Simple RNN with tanh activation functions. Figure 7 shows the architecture 1D CNN and hybrid DNNs models for short-term water demand forecasting examined in this study.

5.3. Implementation of multichannel hybrid models

In hybrid multichannel models, the input data are entered into several channels of 1D CNN with different filter sizes. Finally, they are concatenated together and then entered into RNN layers. In this study, we implemented hybrid multichannel models, CNN-GRU, CNN-LSTM, and CNN-Simple RNN, with 2–5 1D CNN channels according to the architecture shown in Figure 8. We used three convolution layers with the number of filters 64, 128, and 64 in each channel of the 1D CNN. In the first channel, we chose a kernel size of 3; in the next channel, we added two units to the kernel size so that in the five-channel model, we considered a kernel size of 11. Figure 8 shows the architecture of multichannel DNNs for CNN-GRU. For CNN-LSTM and CNN-Simple RNN hybrid multichannel models, its convolution layers are similar to CNN-GRU, and its RNN layers are the same as the layer used in the hybrid single-channel model in Figure 7.

In this study, the weights of the neural networks were modified in an iterative process by the backpropagation algorithm, and using the gradient descent algorithm with a mini-batch size of 256 with the mean absolute error (MAE) loss function. Also, Adam's algorithm (Kingma & Ba 2014) was used to adjust the learning rate, which performed better than other adaptive algorithms. The uniform Xavier Glorot model (Glorot & Bengio 2010) was used for the initial weighting of the parameters, which obtained better answers than other methods. Using Batch Normalization in network layers did not positively affect the results.

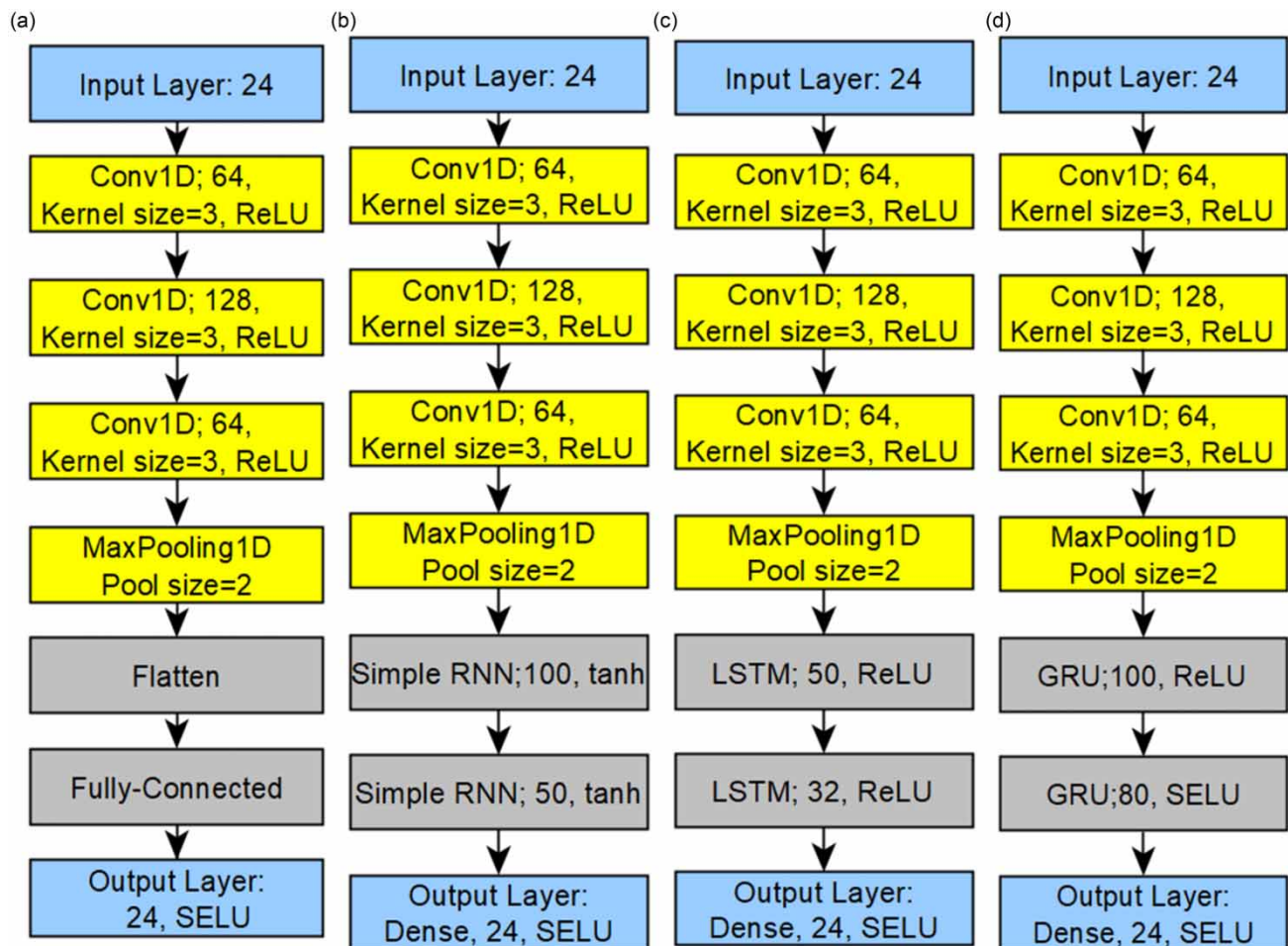


Figure 7 | The architecture of 1D CNN and single-channel hybrid DNN models. (a) 1D CNN, (b) CNN-Simple RNN, (c) CNN-LSTM, and (d) CNN-GRU.

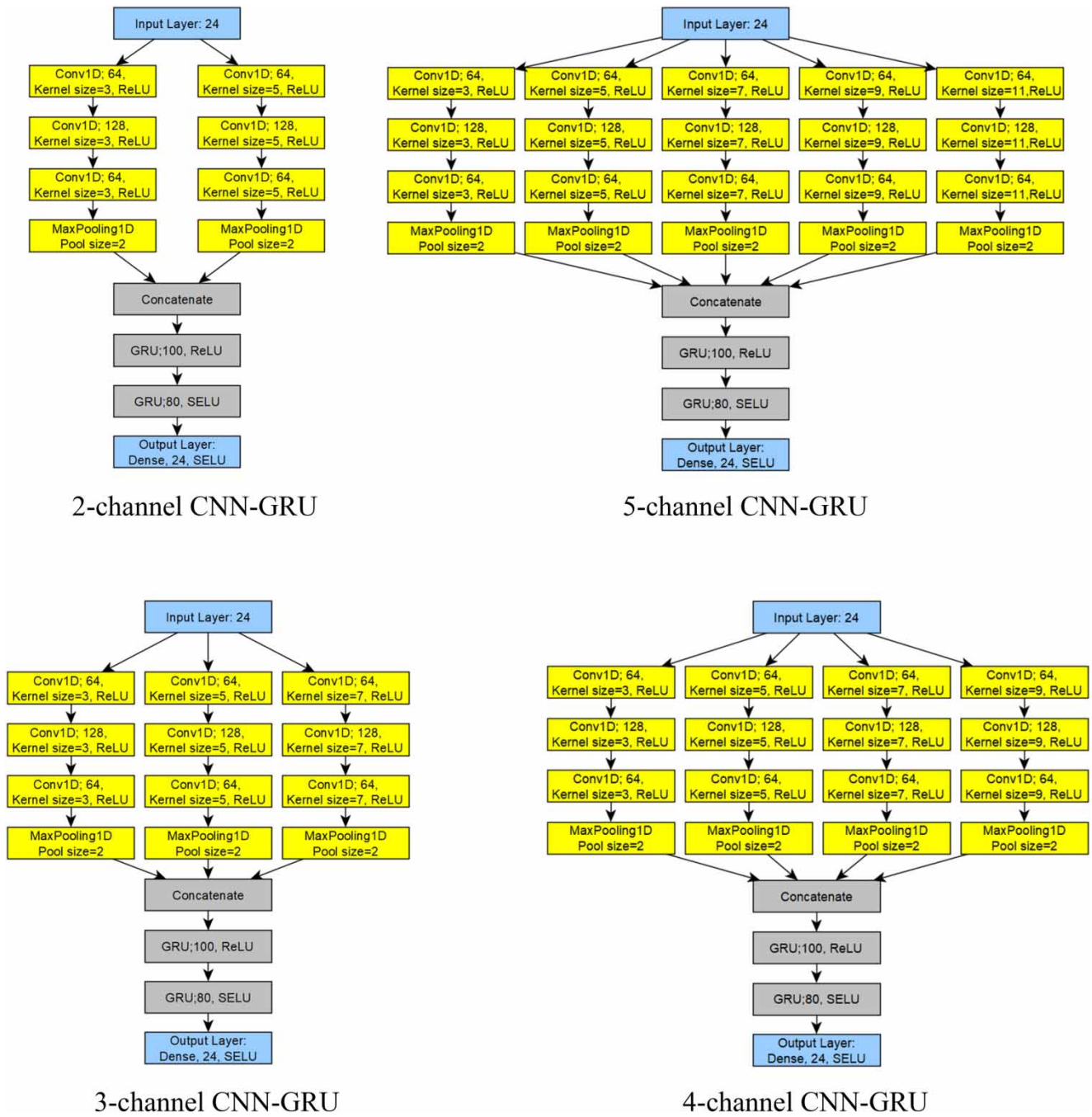


Figure 8 | Architecture of multichannel CNN-GRU.

6. EVALUATION METRICS

The results of the models were evaluated for the test set and training set. The used evaluation metrics include MAE, mean absolute percentage error (MAPE), mean square error (MSE), root mean square error (RMSE), and R^2 score:

$$MAE(y - \hat{y}) = \frac{1}{N} \sum_{i=1}^N |y - \hat{y}|, \tag{4}$$

$$MAPE(\%) = \frac{100}{N} \sum_{i=1}^N \frac{|y - \hat{y}|}{y}, \tag{5}$$

$$MSE(y - \hat{y}) = \frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2, \tag{6}$$

$$RMSE(y - \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2}, \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y - \hat{y})^2}{\sum_{i=1}^N (y - \bar{y})^2} \tag{8}$$

In these equations, y is the real value, \hat{y} is the forecast value, and \bar{y} is the mean of the real values of the samples. The R^2 index indicates the amount of adaptation of the forecast values with the real values. The closer this index is to one, it indicates that the predicted values are closer to real values and the accuracy of the model is higher.

7. RESULTS AND DISCUSSION

In this study, we implemented the 1D CNN model and hybrid DNNs, including CNN-Simple RNN, CNN-LSTM, and CNN-GRU to forecast short-term urban water demand. At first, we used the hybrid models as single-channel ones. The architecture of these models was according to Figure 7. The evaluation results of these models are given in Table 2. The results show that

Table 2 | Evaluation metrics for single-channel hybrid DNNs and 1D CNN for training and test data

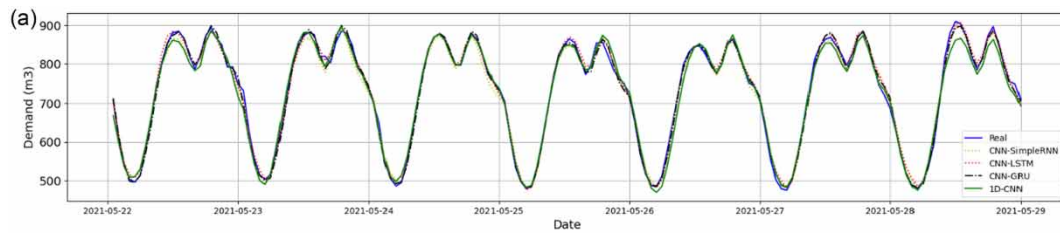
Model	RMSE		MSE		R ²		MAE		MAPE%	
	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training
1D CNN	32.68	24.59	1,069	604.7	0.941	0.855	21.18	14.55	3.13	2.62
CNN-Simple RNN	31.20	23.83	974.0	567.9	0.946	0.863	19.79	13.86	2.92	2.50
CNN-GRU	28.09	22.29	788.8	496.7	0.957	0.881	16.58	12.24	2.51	2.21
CNN-LSTM	30.24	22.70	915.9	515.4	0.950	0.875	19.25	13.00	2.88	2.37

The evaluation metrics of the CNN-GRU model are better than other models, which is why the values of CNN-GRU rows are in bold.

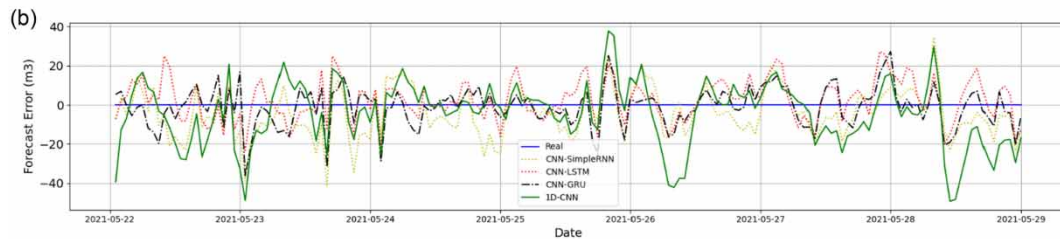
Table 3 | Evaluation metrics in multichannel hybrid DNNs for test and training data

Model	RMSE		MSE		R ²		MAE		MAPE%	
	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training
Two-channel CNN-Simple RNN	28.12	23.21	792.4	538.8	0.956	0.869	16.76	13.30	2.52	2.43
Two-channel CNN-GRU	25.01	20.25	626.1	410.2	0.966	0.901	13.34	10.06	2.03	1.85
Two-channel CNN-LSTM	27.66	21.57	765.2	465.4	0.958	0.888	16.26	11.41	2.47	2.07
Three-channel CNN-Simple RNN	25.13	20.38	634.1	415.3	0.955	0.899	13.71	10.36	2.11	1.90
Three-channel CNN-GRU	23.92	19.64	574.4	385.9	0.959	0.907	12.38	9.53	1.92	1.75
Three-channel CNN-LSTM	25.39	20.88	647.7	436.0	0.954	0.895	14.08	10.84	2.17	1.98
Four-channel CNN-Simple RNN	24.25	19.87	588.1	394.9	0.968	0.905	12.56	9.68	1.90	1.77
Four-channel CNN-GRU	22.98	19.45	528.3	378.4	0.971	0.909	10.70	8.96	1.65	1.64
Four-channel CNN-LSTM	24.66	20.60	608.0	424.4	0.967	0.895	12.77	10.91	1.94	2.04
Five-channel CNN-Simple RNN	24.42	19.53	597.0	381.5	0.967	0.908	12.87	9.16	1.94	1.67
Five-channel CNN-GRU	23.11	18.86	534.5	355.8	0.971	0.914	11.35	8.40	1.72	1.54
Five-channel CNN-LSTM	24.09	19.36	580.9	374.8	0.968	0.910	12.48	9.10	1.89	1.65

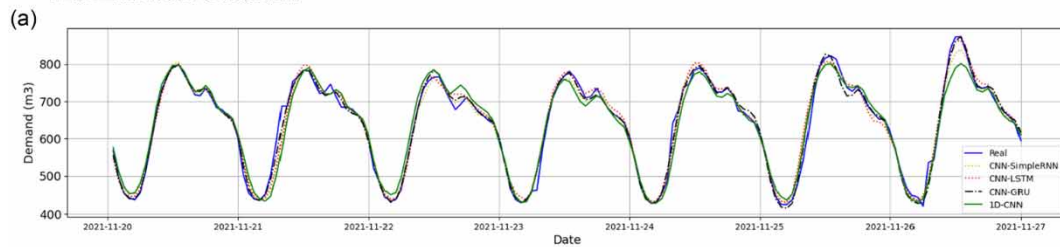
The evaluation metrics of the CNN-GRU model are better than other models, which is why the values of CNN-GRU rows are in bold.



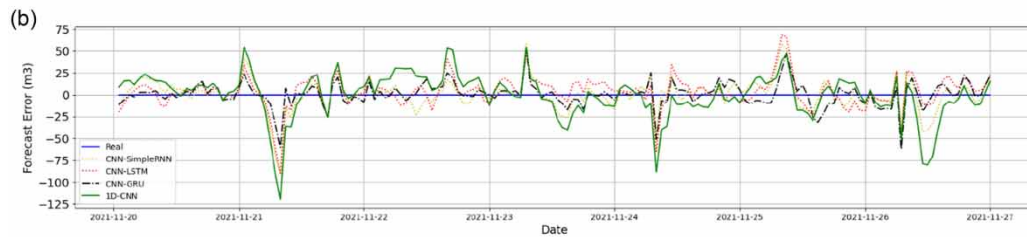
The initial week of test data



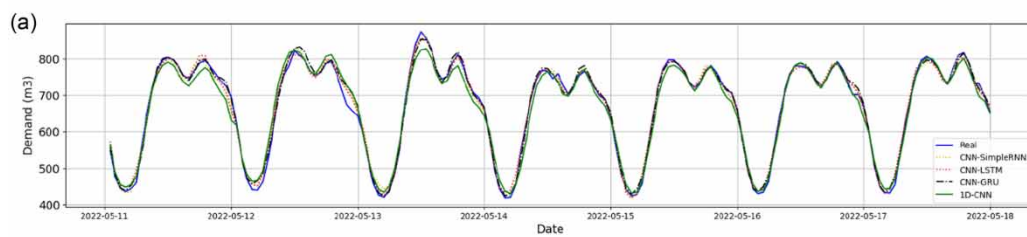
The initial week of test data



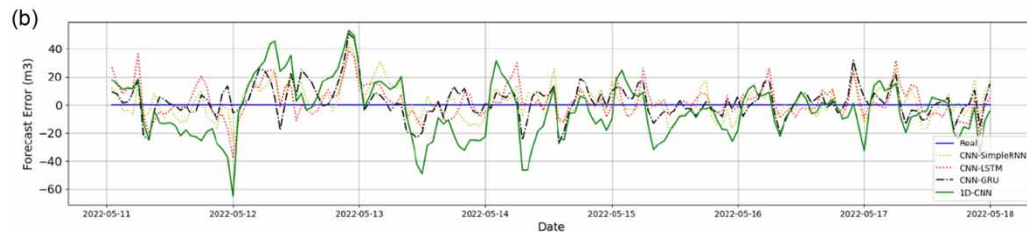
The middle week of test data



The middle week of test data



The end week of test data



The end week of test data

Figure 9 | (a) Forecast and real values in four-channel hybrid models and 1D CNN. (b) The forecast error for each model over time.

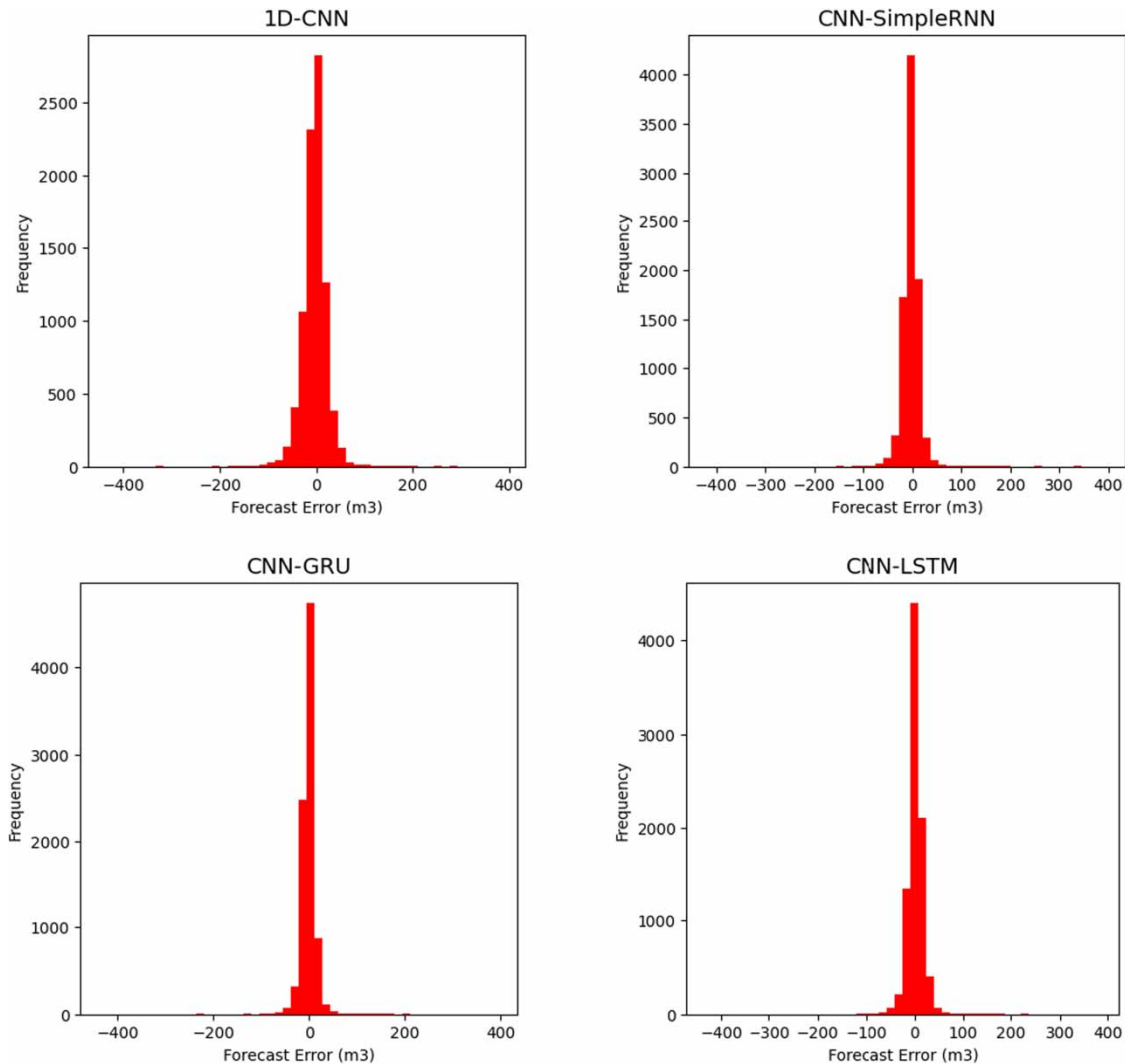


Figure 10 | Histogram of the frequency distribution of the forecasting error in four-channel hybrid models and 1D CNN.

the combination of CNNs and RNNs (hybrid DNNs) improves the forecast results and can increase the accuracy of forecasting water demand. Among the hybrid models, the single-channel CNN-GRU obtained better results than other single-channel models. In the next step, we increased the number of 1D CNN channels in the hybrid models. Increasing the number of 1D CNN channels resulted in better results in all models compared to the single-channel model.

Table 3 shows the evaluation metrics for the hybrid DNNs with two to five CNN channels. As is clear in this table, the evaluation metrics in all the hybrid DNNs improve with the increase in the number of up to four CNN channels. The multi-channel CNN-GRU model with two to five channels has better evaluation metrics than other hybrid models. By increasing the number of CNN channels from four to five, there was no noticeable change in the results; for this reason, we presented the four-channel CNN-GRU model as a proposed model. However, the evaluation metrics on training data in the five-channel models are better than the four-channel models. Figure 8 shows the architecture of the multichannel CNN-GRU. According to this figure, filters of different sizes have been used in the convolution layers on each CNN channel. The variety of filter sizes can extract more diverse features compared to single-channel hybrid models, and ultimately increase the accuracy of forecasting the results. However, with the increase in the number of CNN channels, the training time of the model increases.

Figure 9(a) shows the actual and predicted demand values by all four-channel models and 1D CNN for 1 week from the initial, middle, and end of the test data. Figure 9(b) presents these models' forecast errors over time. Although all the models predict the trend of changes and the minimum and maximum demand values well, hybrid models have much

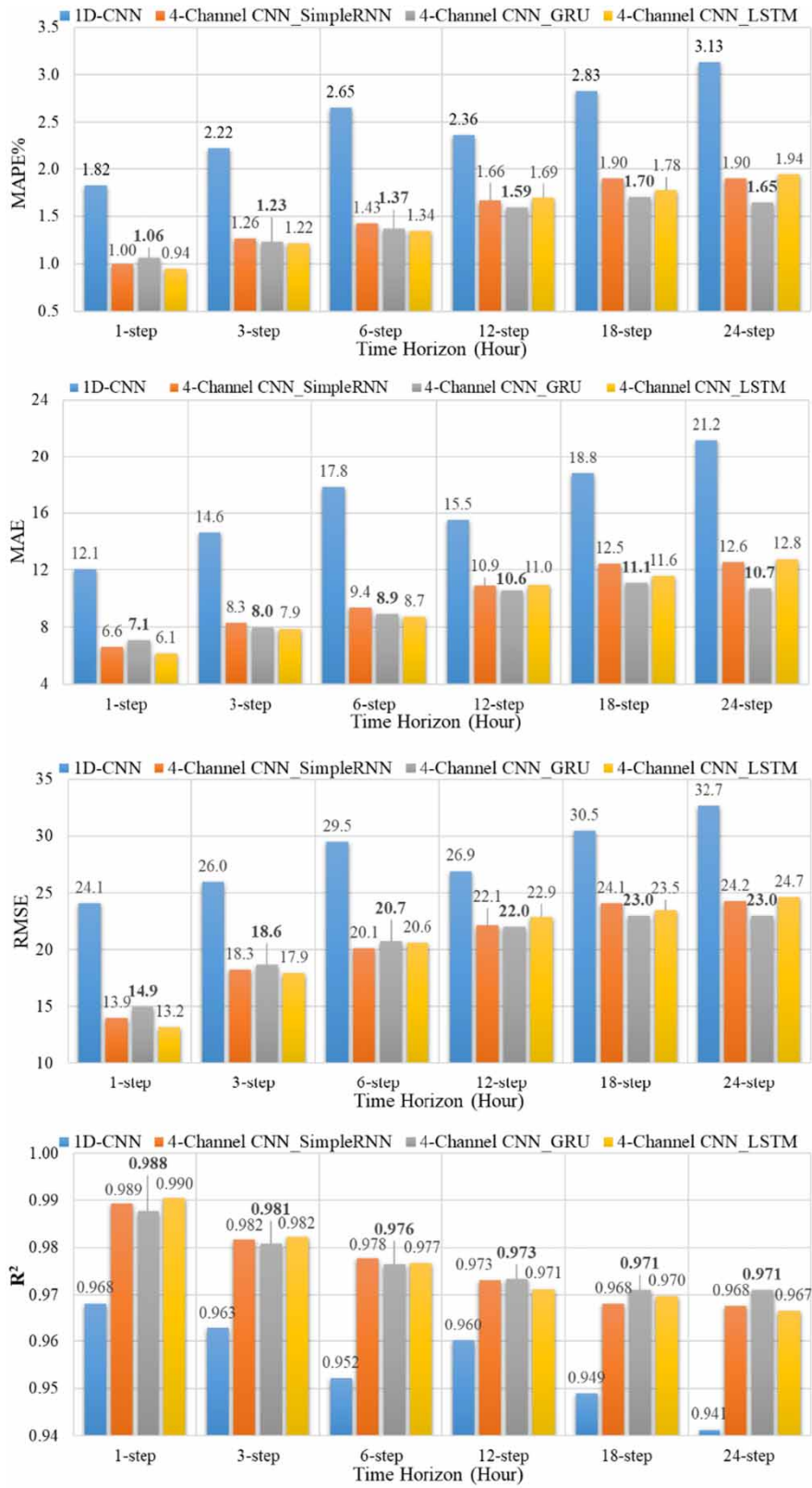


Figure 11 | MAPE, MAE, RMSE, and R² in four-channel hybrid models and 1D CNN for various time horizons.

better accuracy than the 1D CNN model. In the four-channel CNN-GRU, there is a better agreement between the forecast and actual values than in other hybrid models. This issue is evident in Figure 9(b), which shows these models' forecast errors over time.

Figure 10 shows the frequency of forecast error for four-channel hybrid models and 1D CNN for all test data. In the four-channel CNN-GRU model, the frequency of forecast error near zero is much higher than in other models. It shows the superiority of this model over the other two hybrid models and the 1D CNN model.

We examined the effect of the forecast horizon on the results. The evaluation metrics of the four-channel hybrid model and 1D CNN were calculated for the forecast horizons of 1, 3, 6, 12, and 18 h. Figure 11 shows the values MAPE, MAE, RMSE, and R^2 for these models for different forecast horizons. As is evident, the accuracy of the models increases with the decrease in the forecast horizon. Also, the hybrid models have better evaluation metrics on all forecast horizons than the 1D CNN model. In hybrid models up to the 6-h forecast horizon, they have almost similar evaluation metrics, but for horizons of more than 6 h, the four-channel CNN-GRU performs better than the other two models.

8. CONCLUSIONS

To supply adequate water and with the appropriate pressure for consumers in urban WDNs, there is a need for intelligent operation management. One of the requirements of intelligent operation management in WDNs is to forecast short-term water demand at an hourly interval for the next 24 h. In this study, we implemented combined CNN and RNN models to forecast short-term urban water demand. Hybrid models included CNN-Simple RNN, CNN-GRU, and CNN-LSTM. The results showed that the combination of CNNs and RNNs has forecast accuracy much better than the 1D CNN model. Among the hybrid models, the CNN-GRU is superior to other models. On the other hand, increasing the number of CNN channels improved the evaluation metrics, and the four-channel CNN-GRU model had higher accuracy than the other two hybrid models. Thus, we proposed the four-channel CNN-GRU for forecasting water demand. However, it is worth noting that the five-channel models had better evaluation metrics on the training data than the four-channel models. The accuracy of the models increases with the decrease in the forecast horizon. The four-channel CNN-GRU has almost the same accuracy as other hybrid models (CNN-Simple RNN and CNN-LSTM) for forecast horizons of up to 6 h, but it is better than other hybrid models for forecast horizons of more than 6 h. As we mentioned, CNNs can extract features, and RNNs can learn long-term dependencies between data, and combining these two networks can forecast water demand with higher accuracy than 1D CNN. Therefore, hybrid DNNs can be suggested for forecasting other time series issues. During this study, we manually tuned the hyperparameters. However, there are several methods available for automatic hyperparameter tuning, such as random search, grid search, and Bayesian optimization. By utilizing these methods, it is possible to obtain models with better accuracy, although it will take a longer time to train the model. Reducing the training time of the model while maintaining its accuracy is a challenging task. In general, models that predict future values based on time series have the weakness that if one of the factors that affect water demand changes significantly compared to its past, such as the price of water, it can adversely affect the model's results. Therefore, it is important to consider this issue when using such models.

ACKNOWLEDGEMENTS

We are grateful to the Shiraz Water and Wastewater Company for providing the water demand data.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Adamowski, J. & Karapataki, C. 2010 Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* **15**, 729–743.

- Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B. & Sliusarieva, A. 2012 Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* **48**. <https://doi.org/10.1029/2010WR009945>.
- Altunkaynak, A., Özger, M. & Çakmakci, M. 2005 Water consumption prediction of Istanbul city by using fuzzy logic approach. *Water Resources Management* **19**, 641–654.
- Arbués, F., Garcia-Valiñas, M. Á. & Martínez-Espiñeira, R. 2003 Estimation of residential water demand: A state-of-the-art review. *The Journal of Socio-Economics* **32**, 81–102.
- Brentan, B. M., Luvizotto Jr., E., Herrera, M., Izquierdo, J. & Pérez-García, R. 2017 Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics* **309**, 532–541.
- Cho, K., Van Merriënboer, B., Bahdanau, D. & Bengio, Y. 2014 On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.
- De Maria André, D. & Carvalho, J. R. 2014 Spatial determinants of urban residential water demand in Fortaleza, Brazil. *Water Resources Management* **28**, 2401–2414.
- De Souza Groppo, G., Costa, M. A. & Libânio, M. 2019 Predicting water demand: A review of the methods employed and future possibilities. *Water Supply* **19**, 2179–2198.
- Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Alan Roberson, J. 2014 Urban water demand forecasting: Review of methods and models. *Journal of Water Resources Planning and Management* **140**, 146–159.
- Firat, M., Turan, M. E. & Yurdusev, M. A. 2009a Comparative analysis of fuzzy inference systems for water consumption time series prediction. *Journal of Hydrology* **374**, 235–241.
- Firat, M., Yurdusev, M. A. & Turan, M. E. 2009b Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water Resources Management* **23**, 617–632.
- Ghalekhondabi, I., Ardjmand, E., Young, W. A. & Weckman, G. R. 2017 Water demand forecasting: Review of soft computing methods. *Environmental Monitoring and Assessment* **189**, 1–13.
- Ghiassi, M., Zimbra, D. K. & Saidane, H. 2008 Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management* **134**, 138–146.
- Glorot, X. & Bengio, Y. 2010 Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings*, pp. 249–256.
- Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R. & Zhu, X. 2018 Short-term water demand forecast based on deep learning method. *Journal of Water Resources Planning and Management* **144**, 04018076.
- Han, J., Pei, J. & Tong, H. 2022 *Data Mining: Concepts and Techniques*. Morgan Kaufmann, Cambridge, MA, USA.
- Herrera, M., Torgo, L., Izquierdo, J. & Pérez-García, R. 2010 Predictive models for forecasting hourly urban water demand. *Journal of Hydrology* **387**, 141–150.
- Hochreiter, S. & Schmidhuber, J. 1997 Long short-term memory. *Neural Computation* **9**, 1735–1780.
- Kim, G. I. & Jang, B. 2023 Petroleum price prediction with CNN-LSTM and CNN-GRU using Skip-connection. *Mathematics* **11**, 547.
- Kingma, D. P. & Ba, J. 2014 Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M. & Inman, D. J. 2021 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing* **151**, 107398.
- Lazzeri, F. 2020 *Machine Learning for Time Series Forecasting with Python*. John Wiley & Sons, Hoboken, NJ, USA.
- Lu, L., Zhang, C., Cao, K., Deng, T. & Yang, Q. 2022 A multichannel CNN-GRU model for human activity recognition. *IEEE Access* **10**, 66797–66810.
- Lv, S., Wang, L. & Wang, S. 2023 A hybrid neural network model for short-term wind speed forecasting. *Energies* **16**, 1841.
- Mouatadid, S. & Adamowski, J. 2017 Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal* **14**, 630–638.
- Mu, L., Zheng, F., Tao, R., Zhang, Q. & Kapelan, Z. 2020 Hourly and daily urban water demand predictions using a long short-term memory based model. *Journal of Water Resources Planning and Management* **146**, 05020017.
- Namdari, H., Haghighi, A. & Ashrafi, S. M. 2023 Short-term urban water demand forecasting; application of 1D convolutional neural network (1D CNN) in comparison with different deep learning schemes. *Stochastic Environmental Research and Risk Assessment*. DOI:10.1007/s00477-023-02565-3.
- Nauges, C. & Whittington, D. 2010 Estimation of water demand in developing countries: An overview. *The World Bank Research Observer* **25**, 263–294.
- Peña-Guzmán, C., Melgarejo, J. & Prats, D. 2016 Forecasting water demand in residential, commercial, and industrial zones in Bogotá, Colombia, using least-squares support vector machines. *Mathematical Problems in Engineering* **2016**, 5712347.
- Qazi, E. U. H., Almorjan, A. & Zia, T. 2022 A one-dimensional convolutional neural network (1D-CNN) based deep learning system for network intrusion detection. *Applied Sciences* **12**, 7986.
- Ristow, D. C. M., Henning, E., Kalbusch, A. & Petersen, C. E. 2021 Models for forecasting water demand using time series analysis: A case study in Southern Brazil. *Journal of Water, Sanitation and Hygiene for Development* **11**, 231–240.
- Sajjad, M., Khan, Z. A., Ullah, A., Hussain, T., Ullah, W., Lee, M. Y. & Baik, S. W. 2020 A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. *IEEE Access* **8**, 143759–143768.

- Schleich, J. & Hillenbrand, T. 2009 Determinants of residential water demand in Germany. *Ecological Economics* **68**, 1756–1769.
- Shirkoochi, M. G., Doghri, M. & Duchesne, S. 2021 Short-term water demand predictions coupling an artificial neural network model and a genetic algorithm. *Water Supply* **21**, 2374–2386.
- Shuang, Q. & Zhao, R. T. 2021 Water demand prediction using machine learning methods: A case study of the Beijing–Tianjin–Hebei region in China. *Water* **13**, 310.
- Song, H. & Choi, H. 2023 Forecasting stock market indices using the recurrent neural network based hybrid models: CNN-LSTM, GRU-CNN, and ensemble models. *Applied Sciences* **13**, 4644.
- Tiwari, M. K. & Adamowski, J. F. 2015 Medium-term urban water demand forecasting with limited data using an ensemble wavelet-bootstrap machine-learning approach. *Journal of Water Resources Planning and Management* **141**, 04014053.
- Wang, J., Wang, P., Tian, H., Tansey, K., Liu, J. & Quan, W. 2023a A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Computers and Electronics in Agriculture* **206**, 107705.
- Wang, L., Xie, D., Zhou, L. & Zhang, Z. 2023b Application of the hybrid neural network model for energy consumption prediction of office buildings. *Journal of Building Engineering* **72**, 106503.
- Wentz, E. A. & Gober, P. 2007 Determinants of small-area water consumption for the city of Phoenix, Arizona. *Water Resources Management* **21**, 1849–1863.
- Xu, J., Wang, K., Lin, C., Xiao, L., Huang, X. & Zhang, Y. 2021 FM-GRU: A time series prediction method for water quality based on seq2seq framework. *Water* **13**, 1031.
- Yang, X., Jiang, Q., Sun, G. & Tian, Y. 2023 Simulation-data-driven load disaggregation based on multi-channel neural network for industrial and commercial users. *IET Generation, Transmission & Distribution* **17** (7), 1652–1662.
- Zhang, G. P. 2001 An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research* **28**, 1183–1202.

First received 12 August 2023; accepted in revised form 12 February 2024. Available online 28 February 2024