Sustainable Cities and Society 95 (2023) 104610

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Short term water demand forecast modelling using artificial intelligence for smart water management

M Kavya [a], Aneesh Mathew [a,*], Padala Raja Shekar [a], Sarwesh P [b]

[a] Department of Civil Engineering, National Institute of Technology, Tiruchirappalli, 620015, Tamil Nadu, India
[b] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, 632014, Tamil Nadu, India

ARTICLE INFO

Keywords:
Smart water management
Water demand forecasting
Artificial intelligence
LSTM

ABSTRACT

Water is an important resource for life and its existence. Water demand is increasing with increasing economic growth and population, while the water availability is continually depleting making an increasing stress on freshwater resources, necessitating monitoring of water consumption. In addition to controlling the water supply with an efficient water management system, automating the system in terms of both monitoring and operation has received a lot of attention in recent years. Short-term water demand forecast aids in the optimal control of a water supply system and its accurate forecasting helps in reducing operating costs and saving energy. Despite extensive research, the use of demand forecasting for efficient water management has yet to be implemented in India. As a result, the focus of this research is primarily on the model for forecasting short-term water demand using artificial intelligence. A comparative study has been carried out between nine machine learning and deep learning models using the water consumption data over the period from 2020 to 2021 for the city of Hubli in Karnataka. Univariate and multivariate time series forecasting models were considered using the 10-min interval flow meter readings to find the most suitable predictive model. For univariate time series forecasting, only the water consumption was used to predict the water demand, whereas, for the multivariate model, climatic parameters, and calendar inputs (like an hour of the day, holidays, etc.) were considered along with the water consumption data. The results suggest that the deep learning models outperformed the machine learning models, and Long-Short Term Memory (LSTM) model demonstrates the best prediction performance in the two scenarios with a mean absolute error of 0.11 m^3/hr for univariate model and 2.96 m^3/hr for the multivariate model. The best predictive model can be used to forecast the short-term water demand for any region to ensure sustainable water resource management.

1. Introduction

Water is one of the most essential elements for the survival of life on Earth. Every living thing on the planet requires water to survive. However, water shortage is a crucial issue today, and water management is one of the global development challenges. Water demand is increasing with increasing economic growth and population, while the water availability is continually depleting. To have an efficient water management system, the controlled water supply to the consumers has awakened great attention in recent years (Jain et al., 2001; Banihabib and Mirkalaei, 2019). Smart water management is a system that is meant to collect meaningful and actionable data about a water distribution network's flow, pressure, and distribution. The major purpose of the system is to ensure that the infrastructure and energy utilized to carry water are appropriately managed.

Domene and Sauri (2006) investigated the link between urbanization and water use. They considered housing type, garden necessities, household size, availability of swimming pool, income, and consumer behavior toward water conservation methods among the influential variables determining water usage in Barcelona. Kenny et al. (2008) suggested that residential water demand is influenced by a variety of factors, some of which are within water utilities' control (e.g., price, water limits, rebate programs) and others that are not (e.g., climate and weather, demographic features). Current operating conditions, past water demand, and socioeconomic and climatic factors such as relative humidity, air temperature, rainfall, and pressure, all influence water demand in the future (Donkor et al., 2012). Most of the studies suggested the household size as one of the major contributing factors to influence the water demand (Rizaiza, 1991; Reitveld et al., 2000; Basani et al.,

* Corresponding author.
E-mail addresses: aneesh52006@gmail.com, aneesh@nitt.edu (A. Mathew).

https://doi.org/10.1016/j.scs.2023.104610
Received 29 June 2022; Received in revised form 20 March 2023; Accepted 22 April 2023
Available online 23 April 2023

Nomenclature

ANN	Artificial Neural Network
ARIMA	Auto Regressive Integrated Moving Average
AR	Auto Regressive
CNN	Conventional Neural Network
DMA	District Metered Area
DNN	Deep Neural Network
GEP	Gene Expression Programming
KNN	K-Nearest Neighbours
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Square Error
ML	Machine Learning
RF	Random Forest
RMSE	Root Mean Square Error
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SLR	Simple Linear Regression
SVR	Support Vector Regression

2004; Arouna et al., 2010; Shekar and Mathew 2022; Hussein et al., 2016; Villar-Navascués et al., 2018). Various literature determined other factors that influence the consumption rate of water as average price of water supplied, size of property and type of housing, education of household head and family, monthly income of the household, water quality and taste (Billings and Jones 2008; Whittington, 2002; Strand and Walker, 2005; Fan et al., 2013; Alvisi et al., 2007; Aly and Wanakule 2004; Caiado 2010; Ghiassi et al., 2008). Many factors like economy, society, demographic features, environmental conditions (climate, rainfall, temperature), water price, household income, metering, etc. have been proven to influence water consumption in studies (Al-Zahrani and AbMonasar 2015; Gupta et al., 2020; Mathew et al., 2022; Cutore 2008). Many studies have been done about water demand forecasting and the factors affecting the water demand. Hence, focusing on demand modelling and forecasting processes is highly beneficial.

Modeling and forecasting water demand for both short and long-term periods is crucial. Forecasting models are used to develop water demand evaluation and monitoring strategies. The most common type of demand forecasting model is data-driven approaches, such as statistical or machine learning methods. Statistical and machine learning techniques are used in data-driven approaches to find patterns between water consumption and the factors affecting water demand. The fact that statistical models must have a predetermined structure is a big disadvantage, making it difficult to find a single mathematical function that performs effectively with a variety of data sets. Furthermore, statistical models usually fail to deal effectively with complex data relationships, and their forecast accuracy decreases as data volume grows (Cassidy et al., 2020; Box and Jenkins 2015; Anele et al., 2017; Raju et al., 2022; Abdou et al., 2016). Traditional methods based on autoregressive integrated moving averages (ARIMA) have also been used to understand and model urban water demand. Machine learning approaches to forecasting water demand have recently acquired acceptance, and their efficacy in water management has recently gained traction. Advances in deep learning provide further reasons to apply them to the domain of water demand forecasting. Many researchers have suggested and implemented various techniques for water demand forecast and the various techniques to predict short-term water demand primarily include statistical methods, Machine Learning (ML) models, Hybrid models, and Deep Neural Network (DNN) models (Koo et al., 2021). Artificial intelligence approaches can offer novel possibilities for prediction when the underlying physical association cannot be explicitly obtained for sustainable water resource management (Niu and Feng, 2021).

Maidment et al. (1985) constructed a daily municipal water use time series model as a function of rainfall and air temperature. The general methods for water demand time series modeling are autoregressive integrated moving average (ARIMA) model and seasonal autoregressive integrated moving average (SARIMA) model. Linear models have been the focus of research and are often used in practice because they are relatively simpler to understand and implement (Kofinas et al., 2014). Zhou et al. 2002) used the Auto-Regressive (AR) model and Fourier series to anticipate daily water demand in Melbourne, Australia, using weather parameters such as maximum temperature, precipitation, and evapotranspiration as inputs. Wong et al. (2010) looked at the trend, seasonality, weather regression, and holidays in Hong Kong's water usage. The model developed explained 83% of the variance of the six factors considered and was then validated using an independent dataset yielding a R^2 value of 0.76. Because linear regression was very straightforward, it was used in many models. However, it was gradually realized that, because variations in water demand are non-linear in character, linear regression methods are not as accurate as nonlinear regression approaches in predicting them. Major studies in nonlinear regression methods were based on SARIMA and the exponential smoothing model (Tiwari et al., 2016; Tiwari and Adamowski 2013; Bata et al., 2020; Bárdossy et al., 2009; Gokul et al., 2023; Bakker et al., 2014; Bakker et al., 2013; Babel et al., 2011; Sarwesh and Mathew 2022; Adamowski et al., 2012; Mouatadid and Adamowski, 2017; Fattah et al., 2018; Gagliardi et al., 2017).

Brentan et al. (2017) employed a model that integrated Support Vector Machine (SVM) and adaptive fourier series to anticipate water demand and the results were better with RMSE of 1.318 l/s, R^2 value of 0.974, and MAE of 3.45 l/s than when the SVM model (RMSE of 4.767l/s, R^2 value of 0.745, and MAE of 12.91 l/s) was used alone. Chang and Liu (2009) used the Radial Basis Function- Artificial Neural Network (RBF-ANN) model to anticipate water demand. Furthermore, for prediction, Braun et al. (2014) used the Support Vector Regression (SVR) and SARIMA models and found that the SVR model performed better than SARIMA. Candelieri (2017) predicted water use in Milan, Italy, using a combination of SVM and clustering. Ibrahim et al. (2020) forecasted demand for a whole county using two different forecasting techniques of SVR and ARIMA, considering the impact of population increase. The research was conducted on Kuwait's daily water use. When the results were compared, ARIMA had MAPE (1.8) and RMSE (9.4), but SVR had MAPE (0.52) and RMSE (2.59), indicating the difference between the projected and actual water usage.

Few studies have been made using hybrid models where a combination of more than one technique is used. Rangel et al. (2017) used a nonparametric method called Nearest Neighbour (NN) node estimation for predicting the daily consumption pattern. Farias et al. (2018) introduced the Qualitative Multi-Model Predictor Plus (QMMP+) model for anticipating water demand. The model was based on the moving average using the NN classification, an ML model, and a calendar effect based on quantitative and qualitative information. The SARIMA model gave superior results for overall daily water demand forecasting whereas the NN model with a calendar impact for daily patterns was used to distribute hourly water demand and was compared to the ANN model. Pu et al. (2023) proposed a hybrid Wavelet-CNN-LSTM model to predict short term urban water demand. This model outperformed when compared with the other models used in the study like ANN, LSTM, GRUN.

Deep learning has emerged as a notable and promising example of a learning algorithm in recent years. Guo et al. (2018) developed a gated recurrent unit network (GRUN) model to anticipate water demand 15 minutes and 24 hours in the future with a 15-minute time step to study the potential of deep learning in short-term water demand forecasting. The performance of the GRUN model was compared with ANN and SARIMA model. To lessen the cumulative mistake, a correction module was employed and proposed. The correction module improved the performance of ANN and GRUN models and GRUN correction module gave best results with MAE of 3.42 m^3/15min and RMSE of 4.82

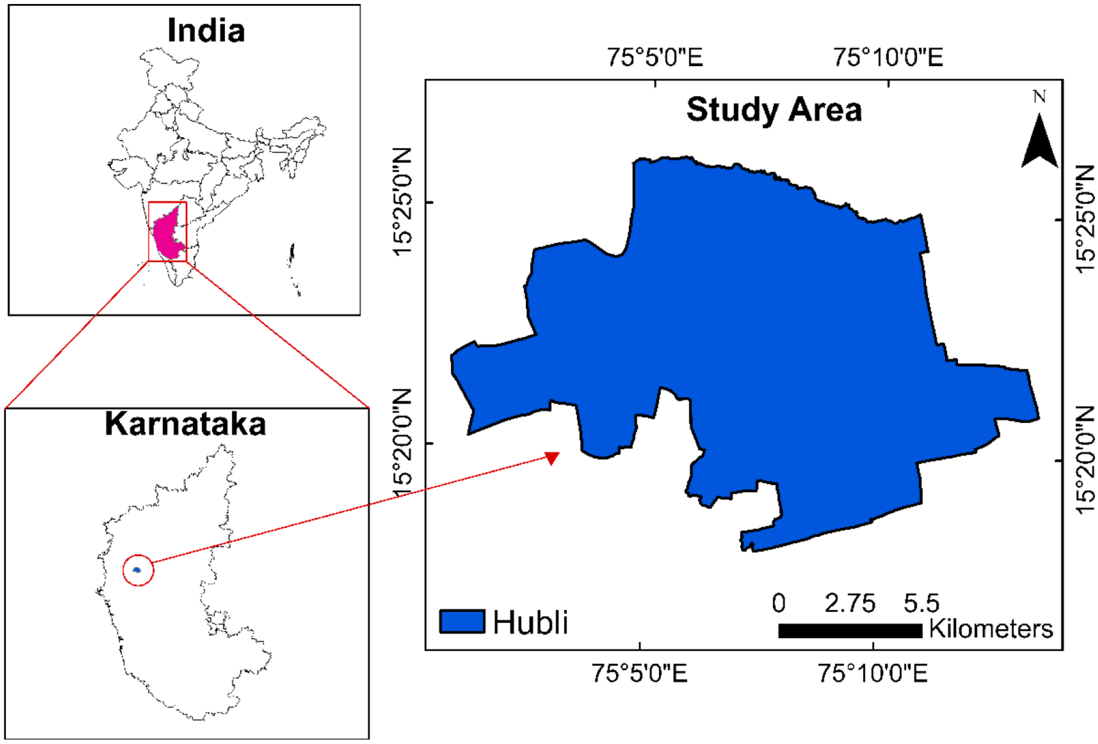


Fig. 1. Map of Hubli city.

m^3/15min when compared to other models. GRUN outperformed ANN and SARIMA models for both 15-minute and 24-hour forecast. According to the findings, the deep learning technology improved the accuracy of water demand forecasting. Koo et al. (2021) investigated short-term water demand forecasting using various models such as ARIMA, RBF-ANN, QMMP+, and long short-term memory (LSTM) on one-hour intervals of real-time water consumption data from smart meters in Yeong Jong Island, Incheon, Korea. It was found that LSTM performed better with RMSE of 56.46 m^3/day, NRMSE of 17.92%, NSE of 0.61, and PCC of 0.79 when compared to the other methods. Zhang et al. (2022) used ARIMA, random forests, LSTM, and neural basis expansion analysis time series (N-BEATS) forecasting models to predict daily and hourly short-term bath water consumption in the shower buildings at China. LSTM model outperformed the other models in both cases of hourly bath water demand (average R^2 of 0.71) and daily bath water demand (MAPE = 5.10%, R^2 = 0.84). Kim et al. (2022) used LSTM approach to predict water demand for four different types of households, namely, detached houses, apartments, restaurant, and elementary school and compared the performance with ARIMA model. LSTM model outperformed ARIMA model for all households with mean R^2 value of 0.89 and RMSE (mean for four cases) of 5.60 m^3. Vangala et al. (2023) predicted future cash withdrawals in automated teller machine (ATM) using various machine learning and deep learning forecasting models like ARIMA, SVR, LSTM, GRU, 1-dimensional convolutional neural network (1D-CNN) and hybrid deep learning techniques like 1D-CNN+LSTM and 1D-CNN+GRU. LSTM yielded better results with lower Symmetric Mean Absolute Percentage Error (SMAPE) values on the test data.

Water demand is increasing with increasing economic growth and population, while water availability is continually depleting, increasing the stress on freshwater resources and necessitating monitoring of water consumption. Accurate water consumption forecasting models are required for fast-growing cities in tropical regions. Water consumption predictions are very useful for the planning of water supply and supporting sustainable city development (Rasifaghihi et al., 2020). The goal of the current work is to address the challenge of precisely predicting short-term water demand for efficient management of water resources. Managing water effectively is essential for guaranteeing its sustainable use because it is a limited and valuable resource. For efficient management of water resources, short-term water demand forecasting is crucial because it enables water providers to plan for water usage, optimize supply, and reduce waste. Traditional approaches to water demand forecasting, however, are frequently constrained by their inability to take into account complex patterns in water consumption, such as abrupt changes in weather or alterations in consumer behavior. This can result in erroneous forecasts, which can then have a negative impact on the environment by causing inefficient use of water resources. Hence the motivation of the present study is to develop a short-term water demand forecasting method based on artificial intelligence that can take into consideration these complex patterns and enhance forecasting accuracy. Thereby, the study aims to contribute to the development of effective smart water management strategies, which can help ensure the sustainable use of water resources and reduce waste. Despite extensive research, the use of demand forecasting for efficient water management has yet to be implemented in India.

As a result, the focus of this research is primarily on the model for forecasting short-term water demand using artificial intelligence, making it a novel study in the region considered. In the present study, nine forecasting models are proposed for urban water consumption using machine learning and deep learning methods. Less research has been done on multivariate models using machine learning and deep learning techniques as compared to univariate models with water consumption as single input. The primary objectives of the present study are to study the factors affecting the water demand in an urban area and to analyze the water consumption patterns during various seasons. Another objective of the study is to forecast the short-term water demand using the actual water consumption and to carry out forecasting using various machine learning and deep learning models and to find the best forecasting model. The necessity of estimating water demand based on actual consumption has been discussed in the present research. The research also discussed the relevance of using advanced machine learning models for water demand forecasting and their benefits. The present study mainly

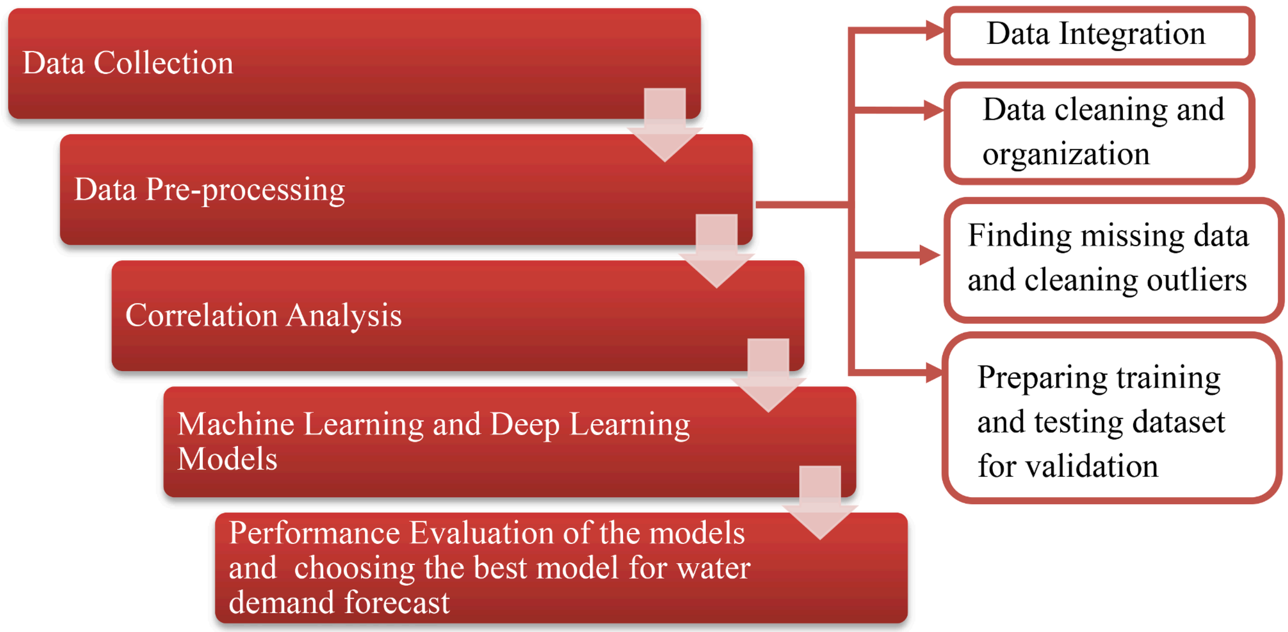


Fig. 2. Methodology adopted for the study.

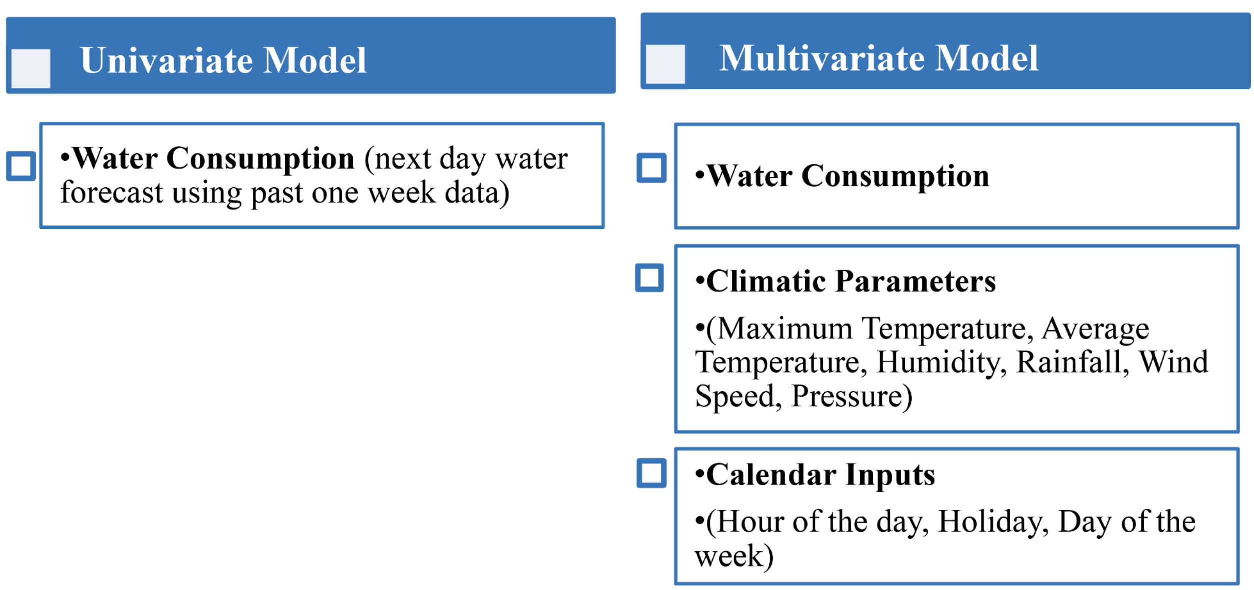


Fig. 3. Input parameters for univariate and multivariate models.

focussed on sustainable water management on urban cities using efficient water demand forecasting. Based on the real characteristics of the water demand in urban centres, it is crucial to carefully select the suitable forecasting models.

2. Study area

Hubli is a city in the state of Karnataka in India. Hubli–Dharwad, the state's second-biggest metropolis by population and size, is also the largest city in North Karnataka. Fig. 1 depicts the map of the city in detail. This study area is selected for this model, where 24 × 7 water supply scheme has been implemented under the smart city project. Hubli has a tropical climate with both wet and dry seasons. From late February to early June, the summers are hot and dry. The monsoon season follows, with mild temperatures and a significant amount of precipitation. From late October to early February, temperatures are mild and there is little rain. Hubli is located at an elevation of 640 meters above sea level. The average annual rainfall is 838 mm. April is the warmest month in Hubli, India, with average high temperatures of 36°C and low temperatures of 22°C. The month with the highest precipitation is July, with an average of 210 mm of rain (Source: https://www.weather-atlas.com/en/india/hubli-climate) August is the coolest month, with an average high of 26°C and a low of 20°C.

3. Data and methodology

Fig. 2 depicts the methodology adopted for this study. The initial phase of the study started with data collection from the study area. After a data pre-processing stage, the acquired data is ready for modelling. Data pre-processing is a 4-stage process involving the following steps:

- Data Integration

Table 1
Climatic parameters considered in the study.

Parameter	Unit
Maximum Temperature	Celsius (°C)
Average Temperature	Celsius (°C)
Humidity	Percentage (%)
Wind speed	Kilometer/hour
Pressure	Millibar
Rainfall	Millimeter

- Data cleaning and organizing
- Check for missing data and outliers
- Preparing training, testing, and validation data sets.

The data pre-processing step is then followed by correlation analysis to find out the correlation between input and output variables. The following nine machine learning and deep learning models will be built for water demand forecasting:

- Linear Regression Model
- Decision Tree Model
- Support Vector Regression Model
- K-Nearest Neighbour Regression Model
- Random Forest Regression Model
- XGBoost Regression Model
- ARIMA Model
- ANN Model
- LSTM Model

Two different types of models, namely univariate and multivariate models will be developed for each of the above-mentioned machine learning models. Fig. 3 depicts the parameters included in each of the models. For univariate time series forecasting, only the water consumption data using the flowmeter readings was used to predict the water demand, whereas, for the multivariate model, climatic parameters, population, and calendar inputs (like an hour of the day, holidays, etc.) were considered along with the water consumption data. After considering the input parameters, the dataset was split as 90% training data and 10% test data, i.e., the model was trained for the dataset from January 2020 to October 2021 and the water demand was predicted for the months of November and December 2021 and the same was validated against the actual consumption pattern available for that duration of 2 months.

The model evaluation for each of these models will be carried out using various metrics and a model comparison will be carried out to choose the best forecasting model for short-term water demand forecast for the study area.

3.1. Data collection

A district metered area (DMA) was chosen for the study from the Hubli region. The 10-minute interval data was collected from the flowmeters for the period of two years from January 2020 to December 2021. The input for climatic parameters like temperature and humidity were collected from the data from Indian Meteorological Department (IMD) (https://mausam.imd.gov.in/). The various parameters for which the data has been collected are given in Table 1.

Apart from the climatic parameters, calendar inputs like an hour of the day, week of the day, etc. were included in the input variables as parameters influencing the water demand. Even though the period of study is considered during COVID as the data was readily available, the dataset considered gives an idea of peak and accurate consumption of the residential area studied and the study duration will not affect the models being tested.

3.2. Data pre-processing

This is one of the very crucial steps before any modelling. The input data must be processed and prepared before it is fed into a machine learning model. This step mainly involves removing inconsistent data. Many a time there are null or missing values in the data. These values must be removed before modelling. Outliers must also be removed from the data before using it in a model. Firstly, the data was grouped such that six 10-min interval readings per hour were clubbed as a single hourly mean value as the study focused on hourly demand forecast of the study region. As the outliers in the data was low, they were identified as per Inter Quantile Range (IQR) (Jones, 2019),

$$\mathrm{IQR} = \mathrm{Q3} - \mathrm{Q1} \tag{1}$$

Where, Q1 is first quartile corresponding 25 percentile, Q3 is the third quartile corresponding 75 percentile. The range considered was (Q1-1.5*IQR, Q3+1.5*IQR).

The outlier points are seen in the data either due to a faulty detection or reading or maybe due to an exceptional event.

To ensure the comparability of their prediction findings, feature scaling was utilised uniformly across all the models studied. The feature scaling method employed was normalisation. The training set is denoted by $D = (X, y)$, where $X = (x_1, x_2, ..., x_n)$ is the n-dimensional explanatory space and y is the dependent variable. The normalisation of x_i is written as

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{2}$$

where xi' is the normalized value of an explanatory variable x for an i^{th} sample; max(x) and min(x) are the maximum and minimum values of x, respectively

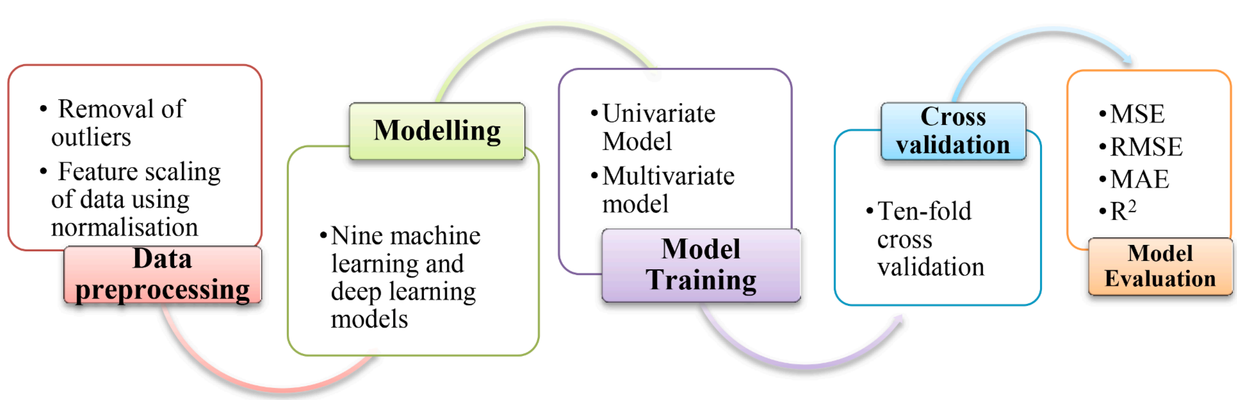


Fig. 4. Data modelling.

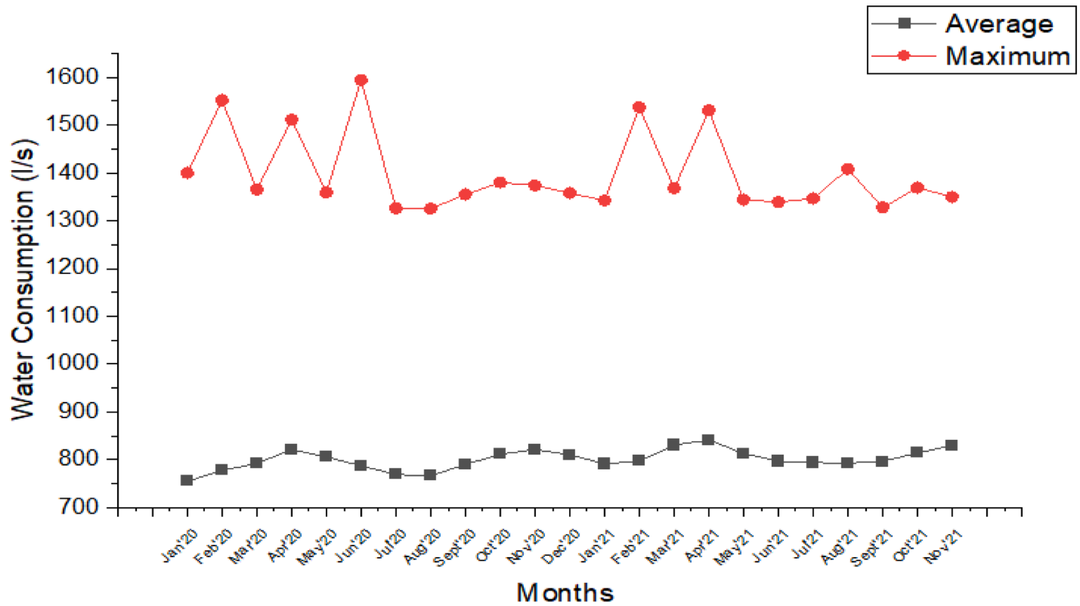


Fig. 5. Average and maximum daily water demand.

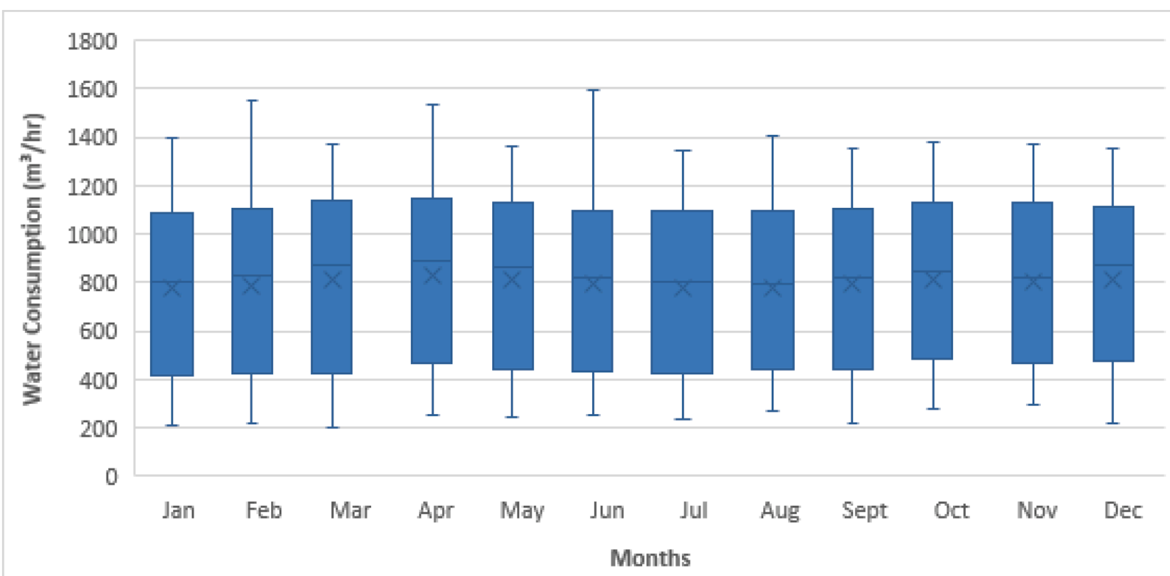


Fig. 6. Box plot showing the monthly variation of water consumption.

3.3. Data modelling

Fig. 4 depicts the process of model preparation. Data pre-processing was the first stage, and it allowed to lower the models' susceptibility to varied data scales. Feature scaling was utilised uniformly across all the models studied to ensure the comparability of their forecast results. The feature scaling method employed was normalisation.

A model is prone to overfitting when fitting a model with the training set and the validating it with the test set, which means it performs well on seen data but poorly on unseen data. A 10-fold cross-validation (CV) was implemented for the training data to solve the issue of overfitting. In this method the training data is divided into ten subgroups at random, each of which represents one-fold. Each model is then trained on nine folds and then validated ten times on the remaining fold, with the validation fold changing each time. The standard deviation and average prediction score are determined using the 10 scores generated by CV (Shaung and Zhao, 2021). Thus, Cross-validation is done for the models to overcome the problem of overfitting.

The model evaluation for each of these models will be carried out using various metrics and a model comparison will be carried out to choose the best forecasting model for short term water demand forecast for the study area.

3.4. Machine learning and deep learning algorithms

In the recent years, with the advent of the field of data science, machine learning and artificial intelligence has been in the limelight and many machine learning algorithms have emerged or have gained more popularity. Supervised machine learning algorithms have been used for the models in this study as per dataset. In the supervised machine learning technique, the mapping function from the input to the output is learnt using the input and output variables of the dataset. The algorithm is trained to make predictions iteratively on the training dataset and give reasonable accurate predictions when new input data is provided. The

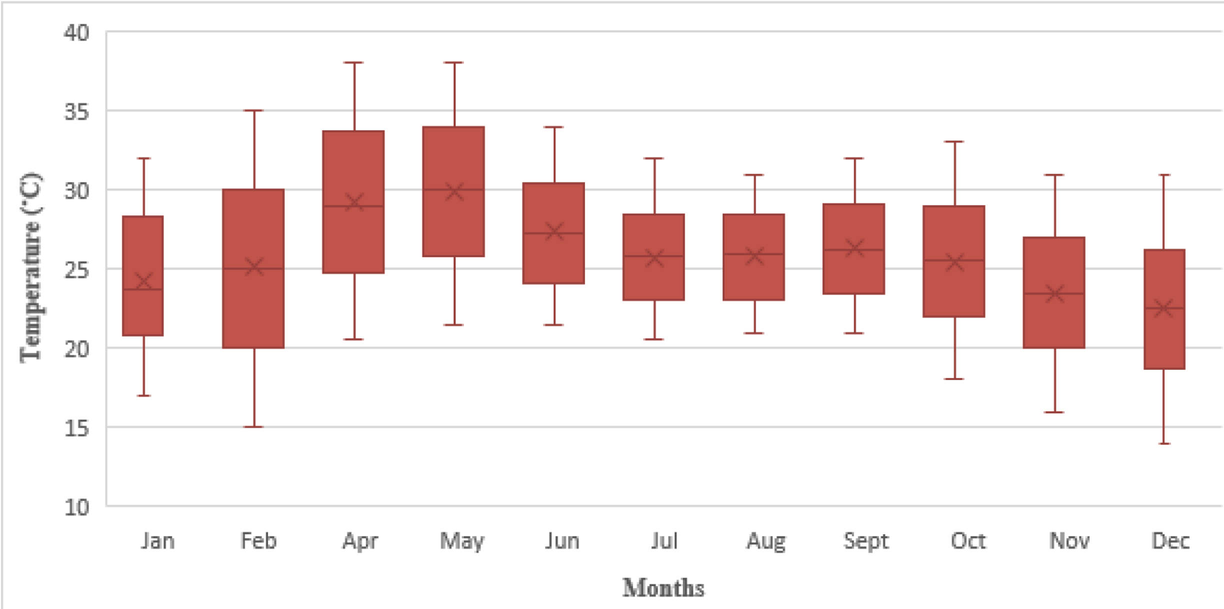


Fig. 7. Monthly variation of temperature.

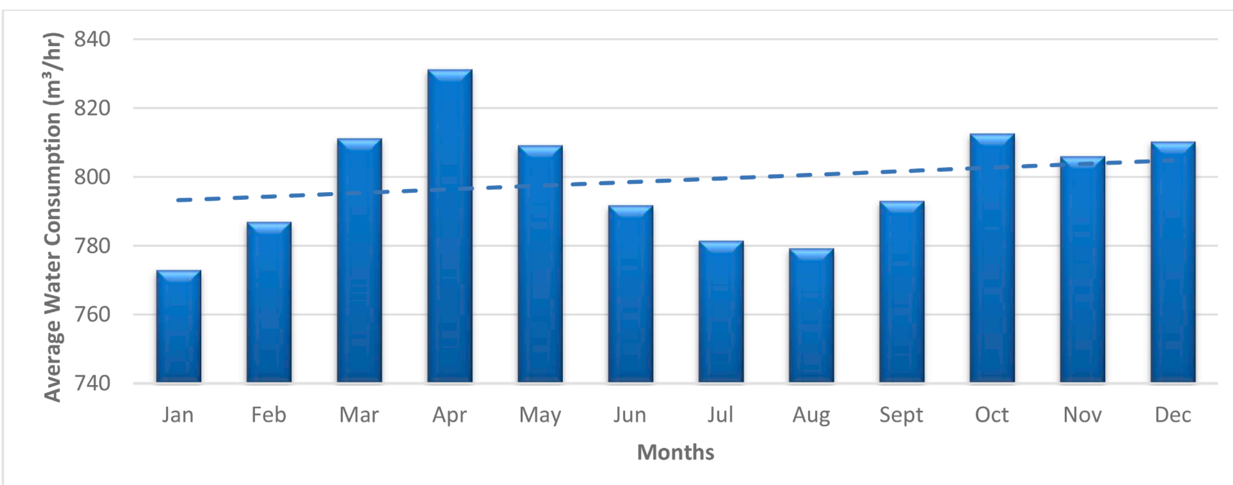


Fig. 8. Monthly variation of average water consumption.

study has been carried out using the libraries readily available in Python. The scikit-learn library in python has been used for creating the models. This library contains various machine learning and deep learning modeling mainly built upon Numpy, Scipy and Matplotlib. The library is also used to perform the cross validation to solve the issue of overfitting. The default parameters available in the scikit-learn library were used for machine learning models like multi linear regression, SVR, DT, Random Forest and XG Boost. For neural network models like ANN, sequential model was created using sequence of layers with input shape as train dataset shape, first input layer with 128 nodes and rectified linear unit (ReLU) activation function, 3 hidden layers with 256 nodes and ReLU activation function, and output layer with linear activation function along with a batch size of 64 with 5 number of epochs. The same parameters for defining neural network were used for LSTM model with mean squared error as loss parameter and adam optimizer.

3.5. Model performance and evaluation

Evaluation is a critical stage in the implementation of any research project. Each model or procedure deployed should be evaluated using one or more metrics to ensure a reliable result. The accuracy of the water demand forecast model is critical to avoid suboptimal system control and to prevent operators from overriding control settings to meet all operational conditions (e.g. to avoid a reservoir to run empty or to overflow) (M. Bakker et. Al., 2013). The performance evaluation of models can be done using various metrics. Based on the various literature studied, the model evaluation metrics used in this study are Root Mean Squared Error, Mean Squared Error, Mean Absolute and R^2 value (Koo et al., 2021; Guo et al., 2018).

The various model evaluation metrics used in the study are as follows:

- **Coefficient of determination (R^2):** R^2 values indicate the goodness of fit between the predicted variables and the test variables and is a reliable indicator for model performance in majority of the cases.

$$R^2 = \left[\frac{1}{M}\left(\frac{\sum_{j=1}^{M}[(Y_j - Y_m)(X_j - X_m)]}{\sigma_x \sigma_y}\right)\right]^2 \quad (4)$$

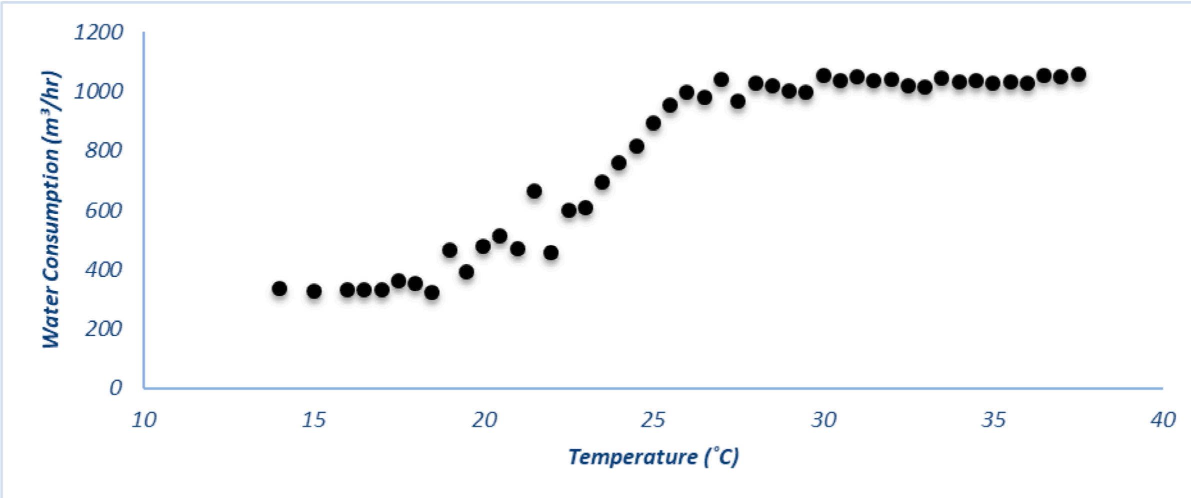


Fig. 9. Average water consumption with respect to temperature.

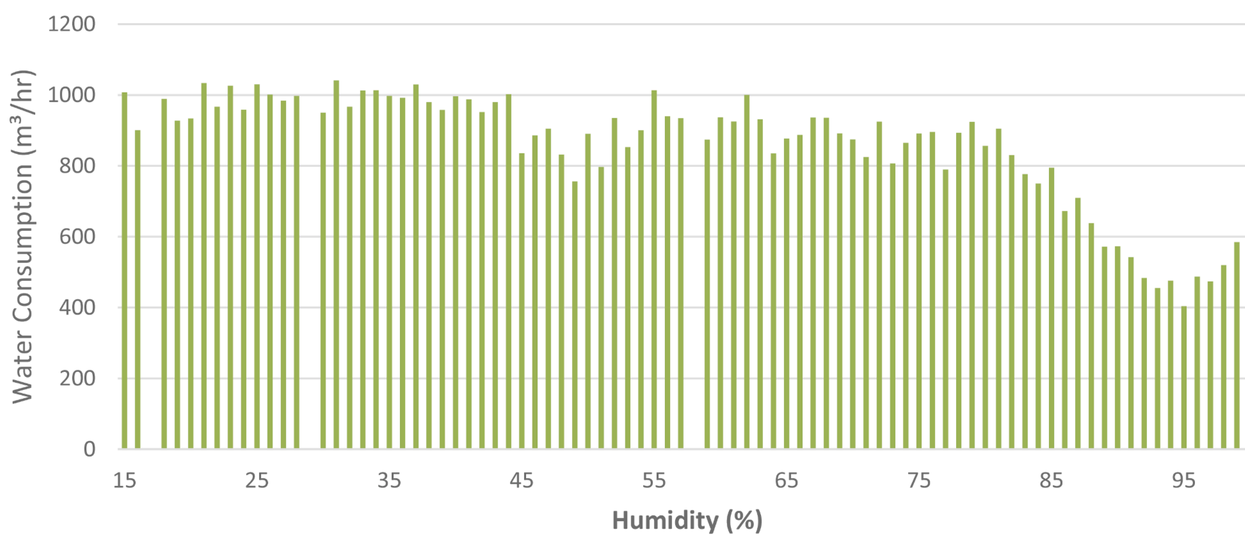


Fig. 10. Average water consumption with respect to humidity.

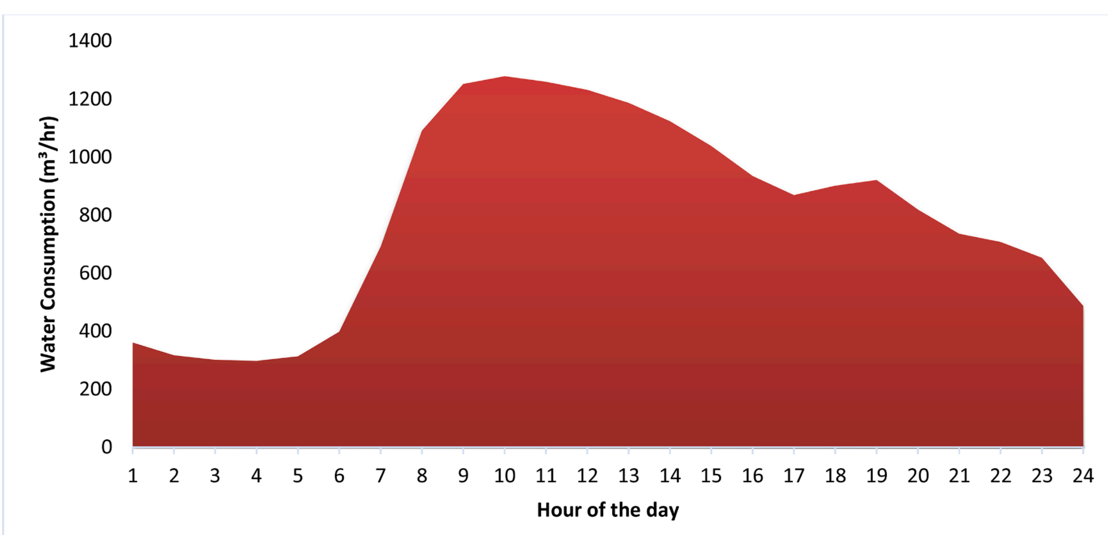


Fig. 11. Hourly average water consumption.

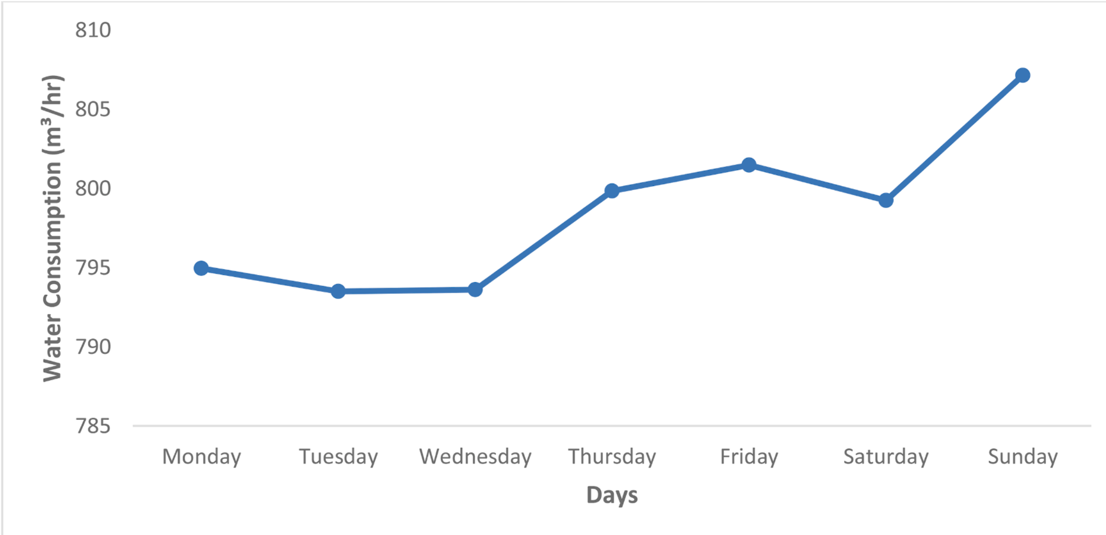


Fig. 12. Weekly water consumption pattern.

where, M - number of observations
σ, – Standard deviation of X and Y respectively
Y_j, - Predicted values and observed values respectively
Y_m, – mean of predicted values and observed values respectively

- **Mean Absolute error (MAE):** The mean absolute error of a model with respect to a test set is the average of all individual prediction errors.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{5}$$

where, n – number of observations y_i – predicted value x_i – actual value

- **Root Mean Squared Error (RMSE):** The RMSE gives information about a model's short-term performance by allowing a term-by-term comparison of the actual difference between the estimated and measured value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (X_i - X_m)^2}{N}} \tag{6}$$

where, N – number of observations,
X_i – Actual observations
X_m –Estimated observations

- **Mean Squared Error (MSE):** MSE tells how close the points are to a regression line.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n} \tag{7}$$

where, n – number of observations y_i – predicted value x_i – actual value

4. Results and discussions

The study primarily focused on analysing the study area's water consumption patterns, researching the factors that influence water demand, and developing models to estimate short-term water demand using various machine learning and deep learning techniques.

4.1. Factors affecting water demand

Much literature has studied about water demand forecasting and the factors affecting the water demand. Many elements, including the economy, society, climate, and environment, have been proven to influence urban water consumption in studies (Shuang and Zhao, 2020). Among all the factors, the major contributing factors that play a vital role are the size of city or town, climatic conditions, cost of water, pressure in the distribution system, economic status, commercial establishments and industries, method of charging, quality of water, sewerage system, and system of supply (Sonowal R., 2017). Based on the studies conducted, water price, household income, type of dwelling, household size, number of taps, garden space, number of children, metering, climate, rainfall, temperature, vegetation area, etc. are the most typically identified factors affecting the water consumption (Kavyashree and Raj, 2020).

4.2. Water consumption pattern of the study area

Based on the 10 min average flow data obtained from the flow meters a clear trend and monthly & daily seasonality was observed. Fig. 5 shows the average and maximum daily demand throughout the study period. The average daily demand was found to be in the range of 750-850 m^3/hr, whereas the maximum demand during the period was between 1300 and 1600 m^3/hr. The total water supplied per day throughout the study period was 11400 m^3. From the Fig. 5, it can be depicted that the average daily demand was almost similar throughout the study period with very less seasonal variation.

Figs. 6 and 7 show the monthly variation of the water consumption and temperature respectively. The average temperature in the region was below 30°C except for the summer when the temperature reached its peak up to 38°C during April and May. The average hourly consumption was stable during the study period in the range of 200 to 1600 m^3/hr with slight increase in consumption during the months with higher temperatures.

Fig. 8 shows the monthly variation of the average consumption of the

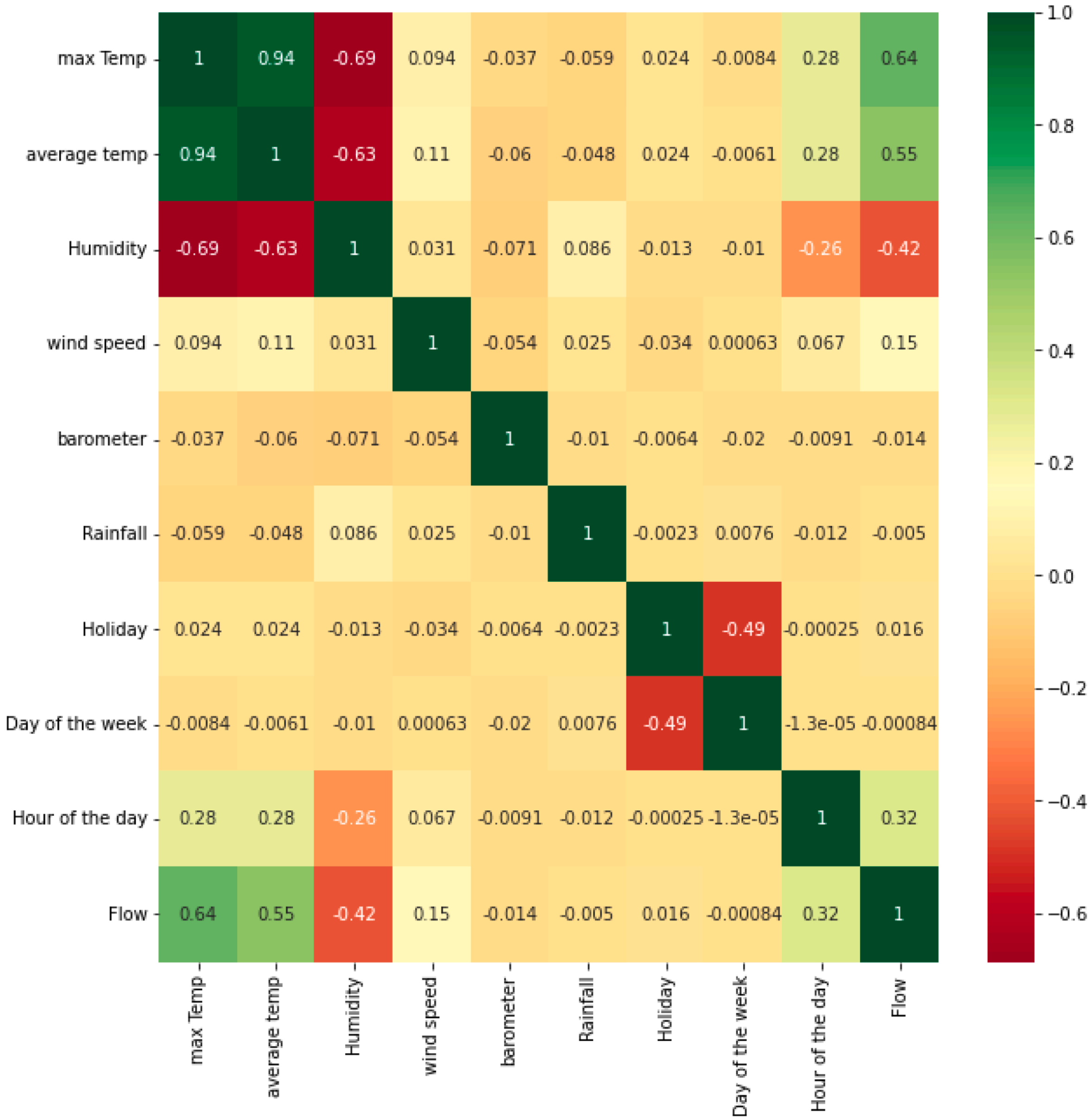


Fig. 13. Correlation heatmap of parameters influencing water consumption.

Table 2
Results obtained from Multi Linear Regression.

Metrics	Univariate Model (MLR1)	Multivariate Model (MLR2)
R^2 (Training data)	0.976	0.965
R^2 (Test data)	0.952	0.56
Cross validation score	0.966	0.964
RMSE (m^3/hr)	2.47	3.65
MAE (m^3/hr)	2.73	5.77
MSE (m^3/hr)	1.2	1.33

study area which varied from 760 to 840 m^3/hr. From the graphs, it was observed that a jump in consumption was observed during the peak summer and peak winter months as obtained from the IMD website. About 8-10% increase in water demand was seen during summers (around April) and peak winters (between October to December). The average demand was found to be in the range of 800 to 840 m^3/hr during the summer months of April and May, whereas the consumption rate is reduced to the range of 760 – 780 m^3/hr during monsoon (June to August). During winters (October to December), the demand was in the range of 810 m^3/hr.

Figs. 9 and 10 show the trend in average water consumption with respect to temperature and humidity respectively. The graph shows that the amount of water consumed increases as the temperature rises. The consumption was very low in the range of 300 m^3/hr during winters when the temperature was around 10-20°C. The highest consumption in the range of 1000 m^3/hr was observed when it was too hot and humid without any rainfall. From the graphs, it can be clearly depicted that when the temperature and humidity are both high, consumption increases; yet, when the humidity is at its highest during rain, consumption drops.

Fig. 11 shows the average consumption of the study area based on the hour of the day. The daily demand varied from 400 m^3/hr to 1300 m^3/hr depending on the hour of the day. The daily consumption pattern implies that demand grows starting at 6 a.m. and peaks in the range of 1300 m^3/hr between 8 AM and 10 AM. It has been noticed that consumption declines during the afternoon and again attains a peak in the range of 800 to 1000 m^3/hr between 5 PM to 7 PM. The demand then gradually decreases during midnight with the lowest consumption rate in the range of 400 m^3/hr.

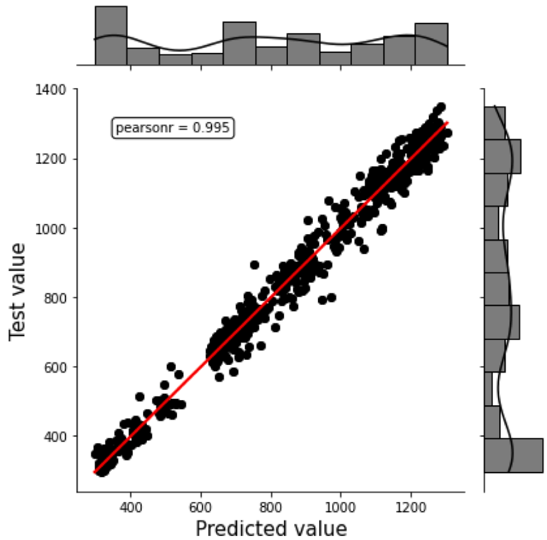


Fig. 14. Joint plot for MLR (a) model 1 (b) model 2.

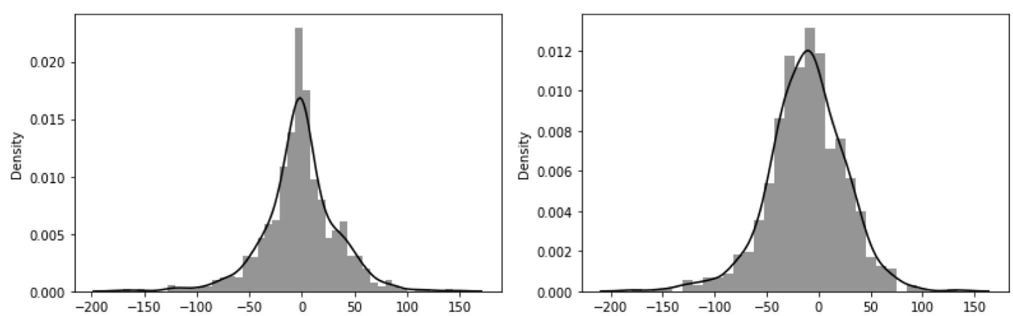


Fig. 15. Distribution plot for MLR (a) model 1 (b) model 2.

Table 3
Results obtained from Decision Tree Model.

Metrics	Univariate model (DT1)	Multivariate model (DT2)
R^2 (Training data)	0.97	0.96
R^2 (Test data)	0.963	0.937
Cross validation score	0.952	0.9503
RMSE (m^3/hr)	4.53	7.89
MAE (m^3/hr)	2.1	1.62
MSE (m^3/hr)	2.86	5.68

Fig. 12 depicts the pattern of water usage with respect to the days of the week. The average consumption on any day was in the range of 790 – 800 m^3/hr while peak consumption of around 810 m^3/hr is observed during weekends especially on Sunday.

After comparing water usage to various climatic and calendar variables, it was observed that consumption in the research area varied according to temperature, hour of the day, and weekday. Throughout the study period, overall daily water consumption followed a similar pattern.

4.3. Correlation analysis

Correlation analysis of the data was carried out to understand the relationship between the input and the output parameters. The analysis was conducted for the climatic and calendar parameters to understand their relationship with water consumption. Fig. 13 shows the correlation heatmap obtained for the parameters.

From the results, it can be inferred that temperature and hour of the day have the maximum impact as they have a positive correlation with water consumption. The maximum temperature and average temperature have positive Pearson coefficient of 0.64 and 0.55 respectively contributing as a major factor affecting the water demand. Whereas humidity has a negative correlation value of -0.42 with the consumption. The negative correlation with humidity can be linked to the decrease in consumption during rainfall when the humidity is high. While considering the calendar parameters, only the hour of the day has a significant positive Pearson correlation factor of 0.32. This can be justified as the consumption was at its peak just before the office hours in the residential area with decline in consumption in the afternoon and again an increase in the demand during the evening time. Other parameters have very small insignificant Pearson correlation coefficients.

The correlation analysis gave an idea about how different input parameters are correlated to the dependent variable. Despite the fact that some parameters have a minor impact on the target variable, all of the parameters have been utilised as input because they do not contribute to the computational complexity of the model.

4.4. Forecasting models

This section describes the results obtained by the forecast models in detail and the evaluation of each model using statistical parameters.

4.4.1. Linear regression model

The results obtained from Multi Linear regression model are given in Table 2. The fitness of both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 14 depicts the joint plot between predicted and actual water

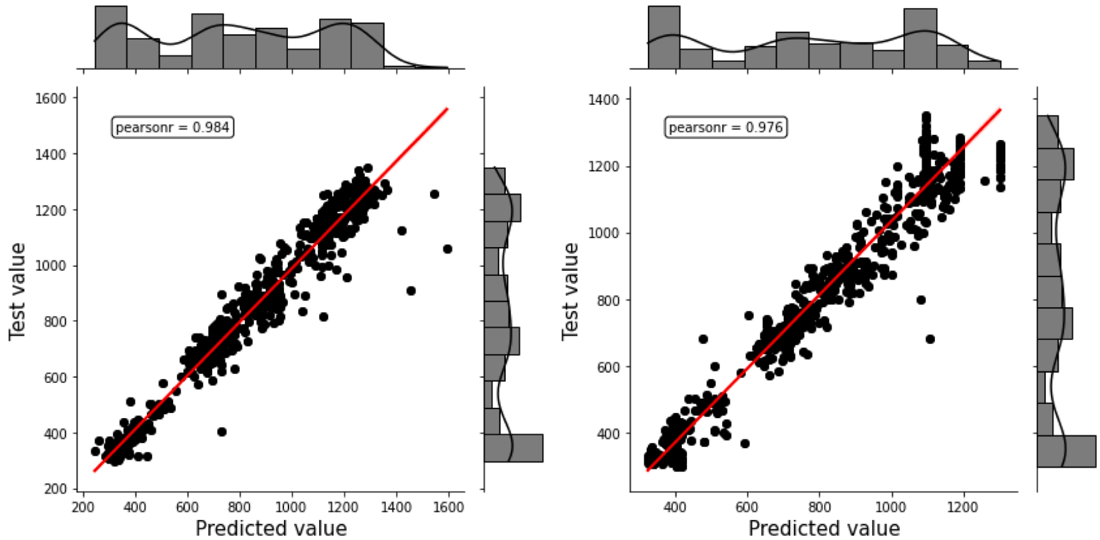


Fig. 16. Joint plot for DT (a) model 1 (b) model 2.

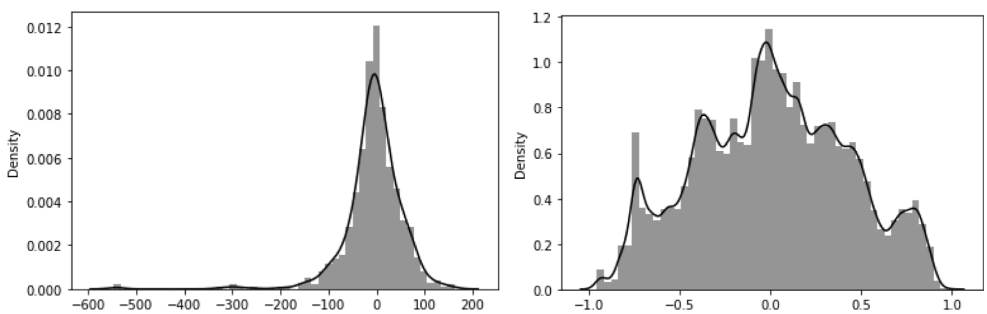


Fig. 17. Distribution plot for DT (a) model 1 (b) model 2.

Table 4
Results obtained from SVR.

Metrics	Univariate model (SVR1)	Multivariate model (SVR2)
R^2 (Training data)	0.964	0.9635
R^2 (Test data)	0.973	0.8845
Cross validation score	0.951	0.881
RMSE (m^3/hr)	3.91	10.65
MAE (m^3/hr)	1.5	2.45
MSE (m^3/hr)	3.31	8.23

consumption values. For MLR1, the R^2 value for testing was 0.956 and the pearson correlation coefficient between predicted and actual values was found to be 0.995 stating a strong relation between predicted and actual values. Whereas, for MLR2, the R^2 value for testing was 0.93 and the pearson correlation coefficient between predicted and actual values was found to be 0.994.

In case of MLR, MLR1 performed better than MLR2 as the error was found to be more in MLR2. The RSME and MAE error obtained in the model MLR1 was 2.47 m^3/hr and 2.73 m^3/hr. The RSME and MAE increased by 47.8% and 111.4% for MLR2 and were found to be 3.65 m^3/hr and 5.77 m^3/hr. Fig. 15 represents distribution plot with kernel density estimator (KDE) function. It shows the distribution of error/residual in the modelling results. The base width of the distribution plot is wider for model 2 when compared to model 1 and the peak obtained for zero error is near 0.015 for both the models indicating that the results obtained could be further improved by using better models.

4.4.2. *Decision tree model*

The results obtained from Decision Tree model are given in Table 3. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 16 depicts the joint plot between predicted and actual water consumption values. For DT1, the R^2 value for testing was 0.96 and the pearson correlation coefficient between predicted and actual values was found to be 0.984 stating a strong relation between predicted and actual values. Whereas, for DT2, the R^2 value for testing was 0.93 and the pearson correlation coefficient between predicted and actual values reduced and was found to be 0.975.

In case of decision tree models, DT1 performed better than DT2 as the error was found to be more than DT2. The RSME and MAE error obtained in the model DT1 was 4.53 m^3/hr and 2.86 m^3/hr. The RSME and MAE increased by 74.2% and 98.6% for DT2 and was found to be 7.89 m^3/hr and 5.68 m^3/hr. Fig. 17 represents distribution plot with KDE. The base width range of model 2 is smaller as compared to model 1. The KDE is not normally distributed for model 2. The model performance of both decision tree models was poor in comparison with MLR models.

4.4.3. *Support vector regression model*

The results obtained from support vector regression model is given in Table 4. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 18 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, SVR1 performed better as the R^2 value for testing was 0.97 and the pearson correlation coefficient between predicted and actual values was found to be 0.993 stating

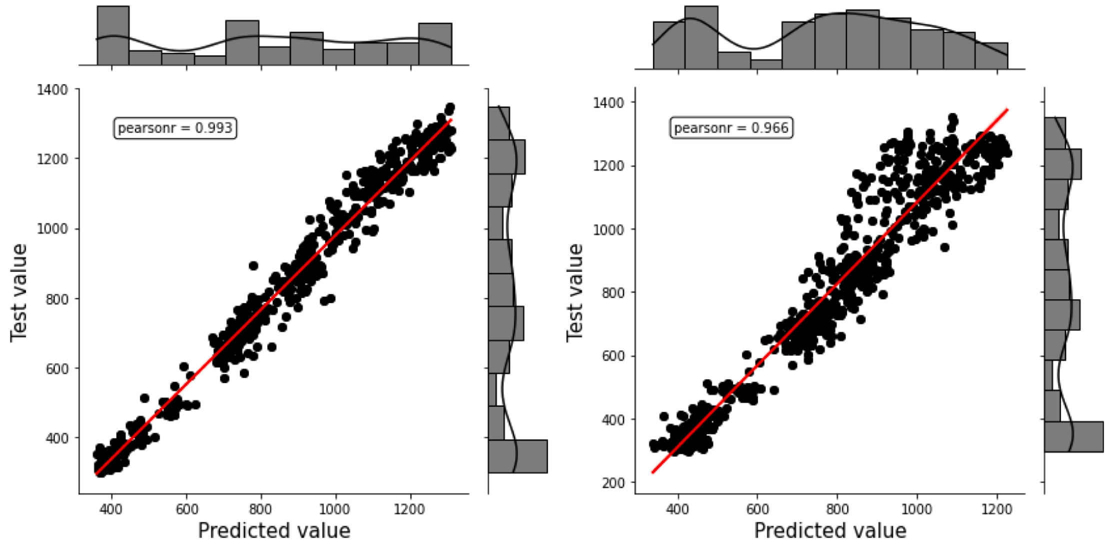


Fig. 18. Joint plot for SVR (a) model 1 (b) model 2.

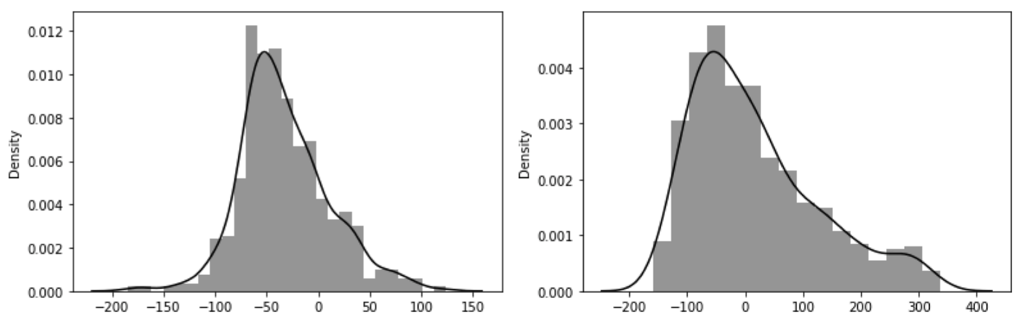


Fig. 19. Distribution plot for SVR (a) model 1 (b) model 2.

Table 5
Results obtained from random forest.

Metrics	Univariate model (RF1)	Multivariate model (RF2)
R^2 (Training data)	0.997	0.987
R^2 (Test data)	0.988	0.968
Cross validation score	0.976	0.962
RMSE (m^3/hr)	2.58	5.59
MAE (m^3/hr)	1.1	3.13
MSE (m^3/hr)	1.82	4.07

a strong relation between predicted and actual values. Whereas, for SVR2, the R^2 value for testing was 0.795 and the pearson correlation coefficient between predicted and actual values reduced and was found to be 0.966.

In case of SVR models, SVR1 performed better than SVR2 as the error was found to be more in SVR2. The RSME and MAE error obtained in the model SVR1 was 3.91 m^3/hr and 3.31 m^3/hr. The RSME and MAE increased by 172.4% and 148.6% for SVR2 and was found to be 10.65 m^3/hr and 8.23 m^3/hr. Fig. 19 represents distribution plot with KDE. The peak of the graph is shifted to the left side for model 1 whereas the error for model 2 is not normally distributed. The model performance of both SVR models was poor in comparison with MLR and DT models.

4.4.4. Random forest regression model

Table 5 gives the results obtained from random forest regression models. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.90.

Fig. 20 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, RF1 performed better as the R^2 value for testing was 0.988 and the pearson correlation coefficient between predicted and actual values was found to be 0.994 stating a strong relation between predicted and actual values. Whereas, for RF2, the R^2 value for testing was 0.959 and the pearson correlation coefficient between predicted and actual values reduced and was found to be 0.989.

In case of random forest models, RF1 performed better than RF2 as the error was found to be more in RF2. The RSME and MAE error obtained in the model RF1 was 2.58 m^3/hr and 1.82 m^3/hr. The RSME and MAE increased by 116.7% and 123.6% for SVR2 and was found to be 5.59 m^3/hr and 4.07 m^3/hr. Fig. 21 represents distribution plot with KDE. The peak for model 1 is slightly near centre, but for model 2, the error is not normally distributed as well as the peak is shifted slightly towards left. The model performance of both RF models was better in comparison with MLR, DT and SVR models as the error produced was less.

4.4.5. XGBoost regression model

Table 6 gives the results obtained from XGBoost regression models. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 22 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, XGB1 performed better as the R^2 value for testing was 0.989 and the pearson correlation coefficient between predicted and actual values was found to be 0.994 stating a strong relation between predicted and actual values. Whereas, for XGB2, the R^2 value for testing was 0.91 and the pearson correlation coefficient between predicted and actual values reduced and was found to be 0.979.

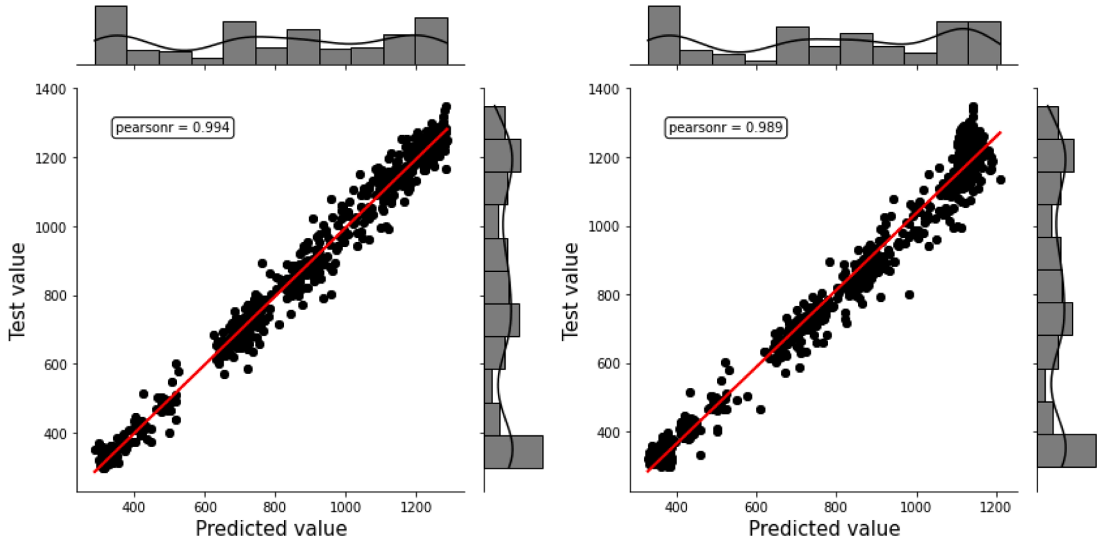


Fig. 20. Joint plot for RF (a) model 1 (b) model 2.

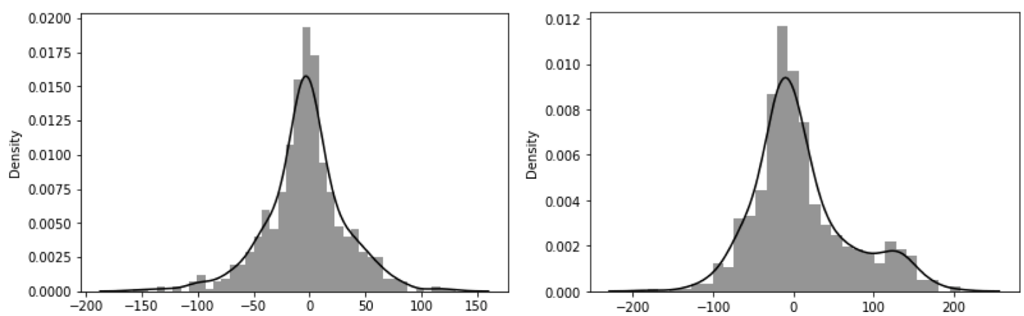


Fig. 21. Distribution plot for RF (a) model 1 (b) model 2.

Table 6
Results obtained from XGBoost.

Metrics	Univariate model (XGB1)	Multivariate model (XGB2)
R^2 (Training data)	0.981	0.9821
R^2 (Test data)	0.989	0.9366
Cross validation score	0.976	0.9759
RMSE (m^3/hr)	2.52	7.89
MAE (m^3/hr)	1.1	1.23
MSE (m^3/hr)	1.79	5.66

In case of XGBoost models, XGB1 performed better than XGB2 as the error was found to be more in XGB2. The RSME and MAE error obtained in the model XGB1 was 2.52 m^3/hr and 1.79 m^3/hr. The RSME and MAE increased by 213.1% and 216.2% for XGB2 and was found to be 7.89 m^3/hr and 5.66 m^3/hr, which is more when compared to MLR2 and RF2. Fig. 23 represents distribution plot with KDE. The base width of model 2 is wider than model 1. The peak of both the models are near centre but the errors for model 2 is not normally distributed. The model performance of XGB1 models was better in comparison with MLR, DT and SVR models as the error produced was less, but RF1 outperformed XGB1. While XGB2 produced more error when compared to RF2 and MLR2.

4.4.6. *K-nearest neighbour model*

Table 7 gives the results obtained from KNN models. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 24 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, KNN1 performed better as the R^2 value for testing was 0.976 and the pearson correlation coefficient between predicted and actual values was found to be 0.988 stating a strong relation between predicted and actual values. Whereas, for KNN2, the R^2 value for testing was 0.915 and the pearson correlation coefficient between predicted and actual values reduced and was found to be 0.972.

In case of KNN models, KNN1 performed better than KNN2 as the error was found to be more KNN2. The RSME and MAE error obtained in the model KNN1 was 3.66 m^3/hr and 2.57 m^3/hr. The RSME and MAE increased by 54.1% and 126.5% for KNN2 and was found to be 5.64 m^3/hr and 5.82 m^3/hr, which is more when compared to MLR2 and RF2. Fig. 25 represents residual distribution plot KDE. The base width of the distribution plot is wider for KNN2 is more than KNN1. The peak obtained for KNN1 is near centre whereas for KNN2, it is slightly shifted towards left. From the graph, it can be depicted that the error is not normally distributed for model 2. The model performance of KNN1 models was better in comparison with DT1 and SVR1 models as the error produced was less, but RF1, and XGB1 outperformed KNN1. While KNN2 produced more error when compared to RF2 and MLR2 but performed better in comparison with DT2 and SVR2.

4.4.7. *ARIMA model*

Table 8 gives the results obtained from ARIMA models. ARIMA model underperformed when compared to all other models in terms of fitness as R^2 values for training and test data was in the range of 0.8 when compared to other models.

Fig. 26 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, ARIMA1 performed better than ARIMA2 as the R^2 value for testing was 0.827 and the

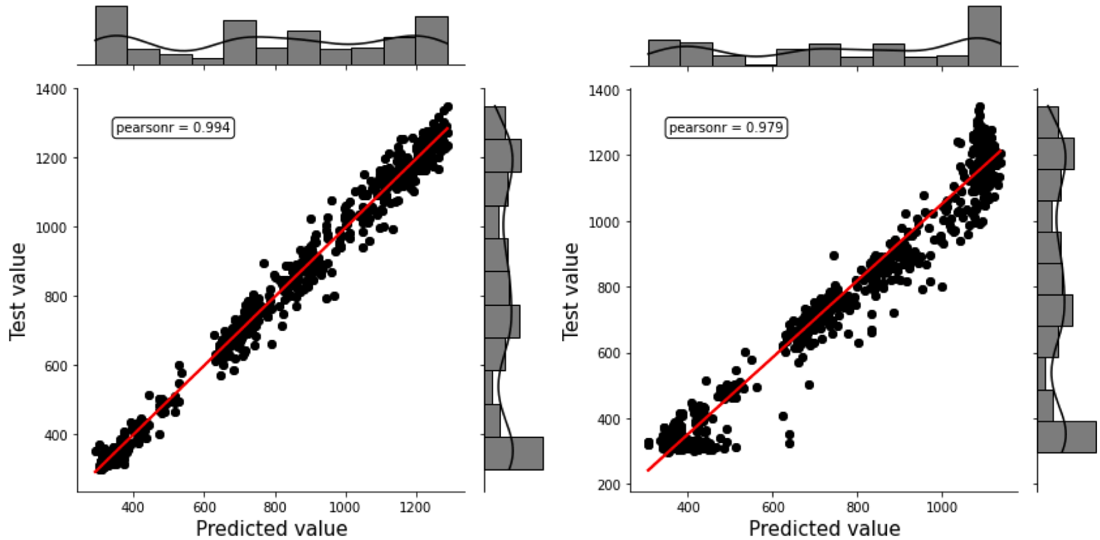


Fig. 22. Joint plot for XGB (a) model 1 (b) model 2.

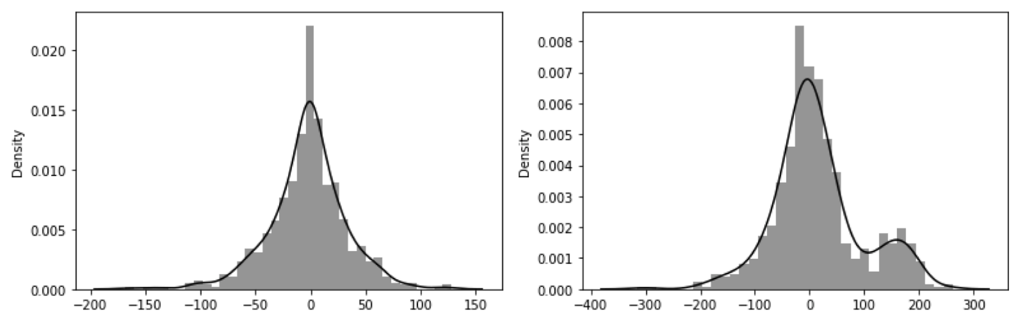


Fig. 23. Distribution plot for XGB (a) model 1 (b) model 2.

Table 7
Results obtained from KNN.

Metrics	Univariate model (KNN1)	Multivariate model (KNN2)
R^2 (Training data)	0.95	0.979
R^2 (Test data)	0.926	0.895
Cross validation score	0.959	0.957
RMSE (m^3/hr)	3.66	5.64
MAE (m^3/hr)	0.98	2.85
MSE (m^3/hr)	2.57	5.82

pearson correlation coefficient between predicted and actual values was found to be 0.856 stating a strong relation between predicted and actual values. Whereas, for ARIMA2, the R^2 value for testing was 0.803 and the pearson correlation coefficient between predicted and actual values reduced and was found to be 0.851.

In case of ARIMA models, ARIMA1 performed better than ARIMA2 as the error was found to be more in ARIMA2. The RSME and MAE error obtained in the model ARIMA1 was 4.65 m^3/hr and 5.62 m^3/hr. The RSME and MAE increased by 89.7% and 42.9% for KNN2 and was found to be 8.82 m^3/hr and 8.03 m^3/hr, which is more when compared to MLR2, RF2, and KNN2. Fig. 27 represents residual distribution plot with KDE. It shows the distribution of error/residual in the modelling results. The errors for both the models are not normally distributed and the peaks are slightly shifted towards left. The model performance of ARIMA models was poor in comparison with other models as the error produced was much higher and R^2 value was also less.

4.4.8. ANN model

Table 9 gives the results obtained from ANN models. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 28 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, ANN1 performed better as the R^2 value for testing was 0.989 and the pearson correlation coefficient between predicted and actual values was found to be 0.995 stating a strong relation between predicted and actual values. Whereas, for ANN2, there was not much difference as the R^2 value for testing was 0.978 and the pearson correlation coefficient between predicted and actual values was found to be 0.994.

In case of ANN models, ANN1 performed better than ANN2 as the error was found to be more in ANN2. The RSME and MAE error obtained in the model ANN1 was 0.06 m^3/hr and 0.25 m^3/hr. The RSME and MAE increased by 200% and 125 % for ANN2 and was found to be 0.193 m^3/hr and 3.48 m^3/hr. In terms of error, ANN models outperformed all other models. Fig. 29 represents distribution plot with KDE. From the Fig., it is clearly visible that the area of graph is more for model 2 as error is distributed more near centre. The model performance of ANN models was much better when compared to other models as the error produced was less.

4.4.9. LSTM Model

Table 10 gives the results obtained from LSTM models. The fitness of the both the models was found to be good with respect to the R^2 values for training and test data as it was above 0.9.

Fig. 30 depicts the joint plot between predicted and actual water consumption values. In terms of fitting if curve, LSTM1 performed better

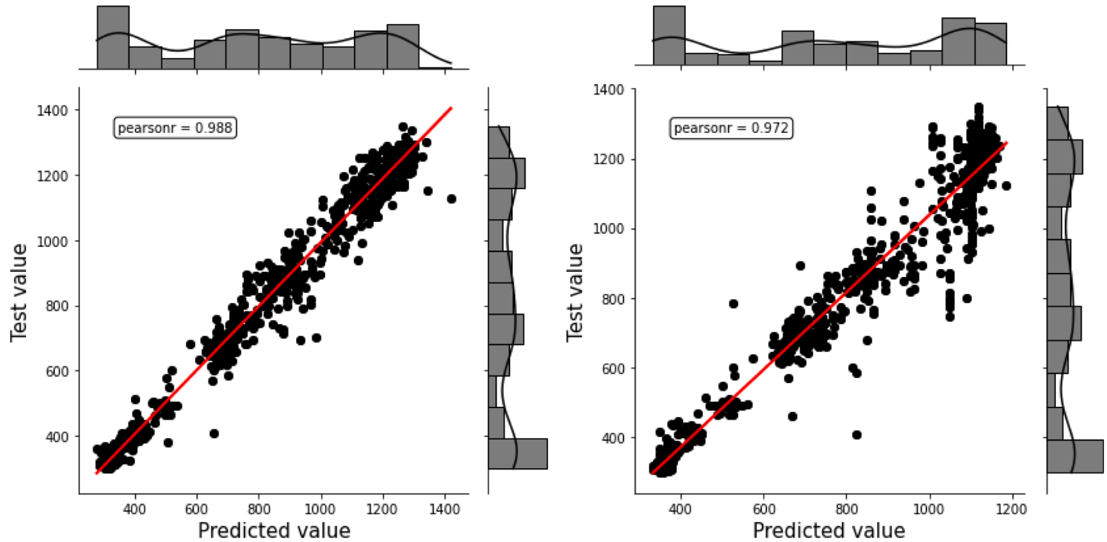


Fig. 24. Joint plot for KNN (a) model 1 (b) model 2.

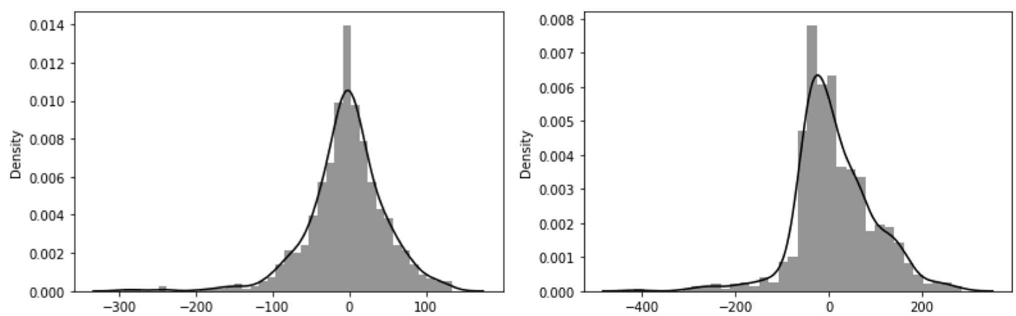


Fig. 25. Distribution plot for KNN (a) model 1 (b) model 2.

Table 8
Results obtained from ARIMA.

Metrics	Univariate model (ARIMA1)	Multivariate model (ARIMA2)
R^2 (Training data)	0.898	0.832
R^2 (Test data)	0.858	0.791
Cross validation score	0.81	0.811
RMSE (m^3/hr)	4.65	8.82
MAE (m^3/hr)	2.3	4.82
MSE (m^3/hr)	5.62	8.03

as the R^2 value for testing was 0.989 and the pearson correlation coefficient between predicted and actual values was found to be 0.995 stating a strong relation between predicted and actual values. Whereas, for LSTM2, there was not much difference as the R^2 value for testing was 0.982 and the pearson correlation coefficient between predicted and actual values rwas found to be 0.994.

In case of LSTM models, both models produced less error as compared to other models. The RSME and MAE error obtained in the model LSTM1 was 0.04 m^3/hr and 0.11 m^3/hr. The RSME and MAE for LSTM2 and was found to be 0.14 m^3/hr and 2.96 m^3/hr. In terms of error, LSTM models outperformed all other models. Fig. 31 represents distribution plot with KDE. The peaks for both the models are near centre and the errors seem to be normally distributed for both models. The model performance of LSTM models was found to be the best suitable model when compared to other models as the error produced was less.

4.5. Comparison of models

4.5.1. Univariate model

The fitness of the models was compared based on R^2 values for training and testing dataset. The models were compared using the statistical parameters of MAE, RMSE and MSE. Table 11 shows the results obtained from various models.

It can be depicted from the results that the fitness was good for all the models. ARIMA model produced the lowest R^2 of 0.898 and 0.858 for the training and test dataset, whereas Random Forest and XGBoost performed satisfactorily with R^2 above 0.98 for both the training and test dataset. ANN model performed better than all other regression models with R^2 values of 0.979 and 0.982 for the training and test dataset. The LSTM model performed better when compared to other models with R^2 values of 0.992 and 0.989 for the training and test dataset.

Among other models, ARIMA produced the maximum error of 5.62 m^3/hr, whereas XGBoost and Random Forest models were found to be satisfactory with mean absolute error of 1.79 and 1.82 m^3/hr respectively. LSTM model outperformed the other models with the least mean absolute error of 0.11 m^3/hr and RMSE value of 0.04 m^3/hr.

4.5.2. Multivariate model

The fitness of the models was compared based on R^2 values for training and testing dataset. The models were compared using the statistical parameters of MAE, RMSE and MSE. Table 12 shows the results obtained from various models.

The comparison of R^2 values during the training and testing shows a clear picture of model performance. It can be depicted from the results that the fitness was good for all the models. ARIMA model produced the

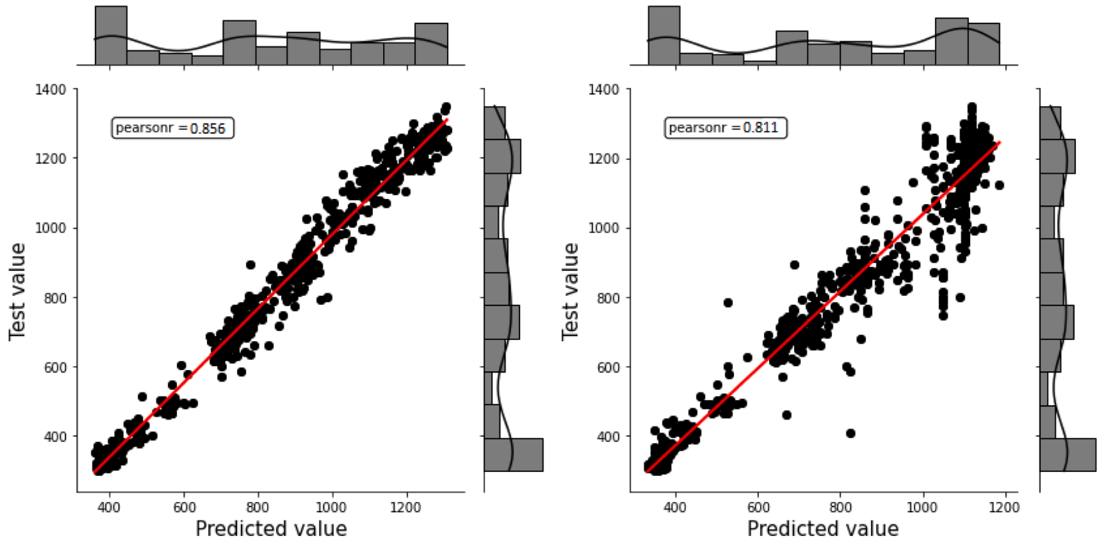


Fig. 26. Joint plot for ARIMA (a) model 1 (b) model 2.

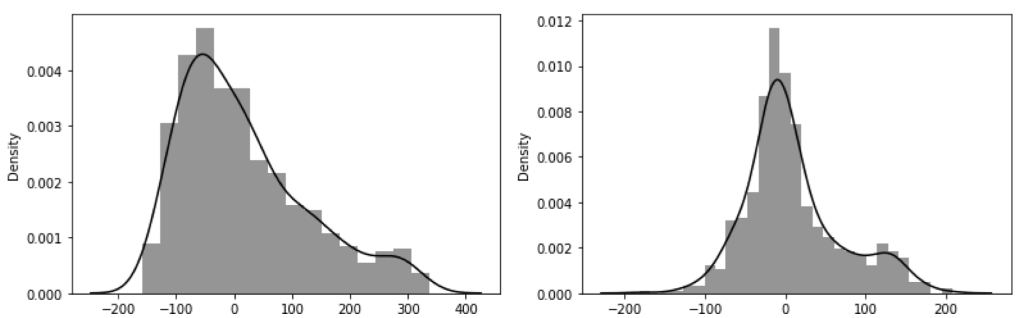


Fig. 27. Distribution plot for ARIMA (a) model 1 (b) model 2.

Table 9
Results obtained from ANN.

Metrics	Univariate model (ANN1)	Multivariate model (ANN2)
R^2 (Training data)	0.979	0.978
R^2 (Test data)	0.982	0.981
Cross validation score	0.972	0.965
RMSE (m^3/hr)	0.06	0.193
MAE (m^3/hr)	0.56	1.93
MSE (m^3/hr)	0.25	3.48

lowest R^2 of 0.832 and 0.791 for the training and test dataset, whereas Random Forest and XGBoost performed satisfactorily with R^2 above 0.96 for both training and test dataset. ANN model performed better than all other models, but the LSTM model outperformed when compared to other models with R^2 value of 0.991 and 0.988 for the training and test dataset.

Among other models, SVR produced the maximum error of 8.23 m^3/hr followed by the ARIMA model with mean absolute error of 8.03 m^3/hr, whereas the Random Forest model was found to be satisfactory with mean absolute error of 4.07 m^3/hr. ANN model performed better than all regression models, but LSTM model outperformed the other models with the least mean absolute error of 2.96 m^3/hr and RMSE value of 0.14 m^3/hr.

4.5.3. Comparison of predicted and actual consumption on test data

After evaluating the models based on the statistical parameters, it was found that LSTM performed better among all the models. So, the LSTM model was used to compare the actual and predicted water demand of the test dataset. Fig. 32 (a) shows the graph of comparison between predicted and actual values for the test dataset. Fig. 32 (b) shows the graph of random samples selected from the test dataset to understand the difference in error between predicted and actual values more clearly. The pattern of predicted water consumption matched with actual water consumption.

4.5.4. Predicting future water demand

As the LSTM model outperformed among all the models, the same model was chosen for forecasting water demand for the period of 2 months. For this, the model was trained for the dataset from January 2020 to October 2021 and the water demand was predicted for the months of November and December 2021. Fig. 33 shows the trend of forecasted water demand. The forecasted values were then compared with the actual consumption reading from the flowmeter. It was found that the demand matched the trend and forecasting was found to be satisfactory with the mean absolute error of 1.23 m^3/hr.

5. Discussions

The best model was found out be LSTM based on maximum R^2 value and minimal error. With the help of this model, operation and storage efficiency can be improved in the Hubli region by predicting the future demand. The forecasted demand can be used to optimize the cost of distribution by managing the supply and storage based on the predicted demand and can help to minimize the losses in the distribution system. This short-term water demand forecast can be helpful to monitor increased water supply and storage required to fulfil future demands. It will be enhancing the capabilities of the treatment plant and water

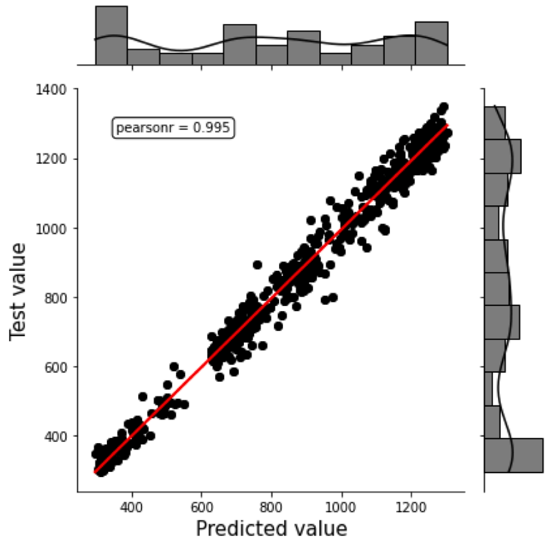


Fig. 28. Joint plot for ANN (a) model 1 (b) model 2.

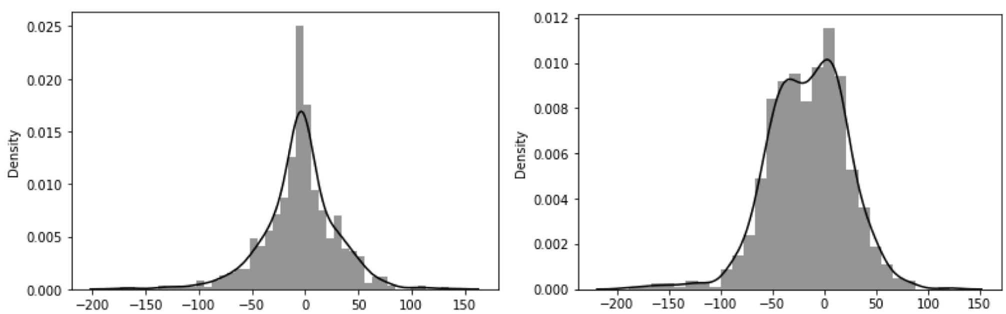


Fig. 29. Distribution plot for ANN (a) model 1 (b) model 2.

Table 10
Results obtained from LSTM.

Metrics	Univariate model (LSTM1)	Multivariate model (LSTM2)
R^2 (Training data)	0.992	0.991
R^2 (Test data)	0.989	0.988
Cross validation score	0.981	0.979
RMSE (m^3/hr)	0.04	0.14
MAE (m^3/hr)	0.1	1.4
MSE (m^3/hr)	0.11	2.96

supply system to handle peak consumption as the predicted demand is available. As the future demand is available, it can be analysed, and water demand can be addressed effectively in case of short-term supply shortage or emergency demand caused by unexpected situations. Both the univariate and multivariate models produced comparable results for LSTM model and suitable model among the two can be used depending on the use and availability of the data. Thus, the LSTM forecasting model can be effectively used to forecast water demand to regulate water pressure, schedule pumping operations, monitor system maintenance, and reduce leakage in the region. Using this short-term water demand forecasting model, the operation and management of an existing water supply system can be managed efficiently and effectively by forecasting the demand based on actual consumption and taking necessary actions as per the demand ensuring smart water management.

6. Conclusions

Analysis of water consumption patterns was done for the city of Hubli for a period from January 2020 to December 2021. Water demand forecasting models were based on machine learning and deep learning techniques.

The average daily demand of the study area was found to be in the range of 750-850 m^3/hr, whereas the maximum demand during the period was between 1300-1600 m^3/hr throughout the study period. The total water supplied per day throughout the study period was 11400 m^3. The average daily demand of Hubli was almost similar throughout the study period with very less seasonal variation. The average hourly consumption was stable during the study period in the range of 400 to 1200 m^3/hr with slight increase in consumption during the months with higher temperature. It was observed that a jump in consumption was observed during the peak summer and peak winter months as obtained from the IMD website. Around 8-10% increase in water demand was seen during summers (around April) and peak winters (between October to December) The daily demand varied from 400 m^3/hr to 1300 m^3/hr depending on the hour of the day. The daily consumption pattern implied that demand grows starting at 6 a.m. and peaks between 8 and 10 a.m., then declines during the afternoon and again attains a peak between 5 to 7 p.m. The demand then gradually decreases with lowest consumption rate during midnight. The water consumption was found to be high during weekends as compared to weekdays. It was found that the selection of input parameters has a very crucial role in modelling. Correlation analysis of the data was carried out to understand the relationship between the climatic and calendar parameters with water consumption. The maximum temperature and average temperature have positive Pearson coefficient of 0.64 and 0.55 respectively contributing as a major factor affecting the water demand. Humidity indicated a negative correlation value of -0.42 with the consumption,

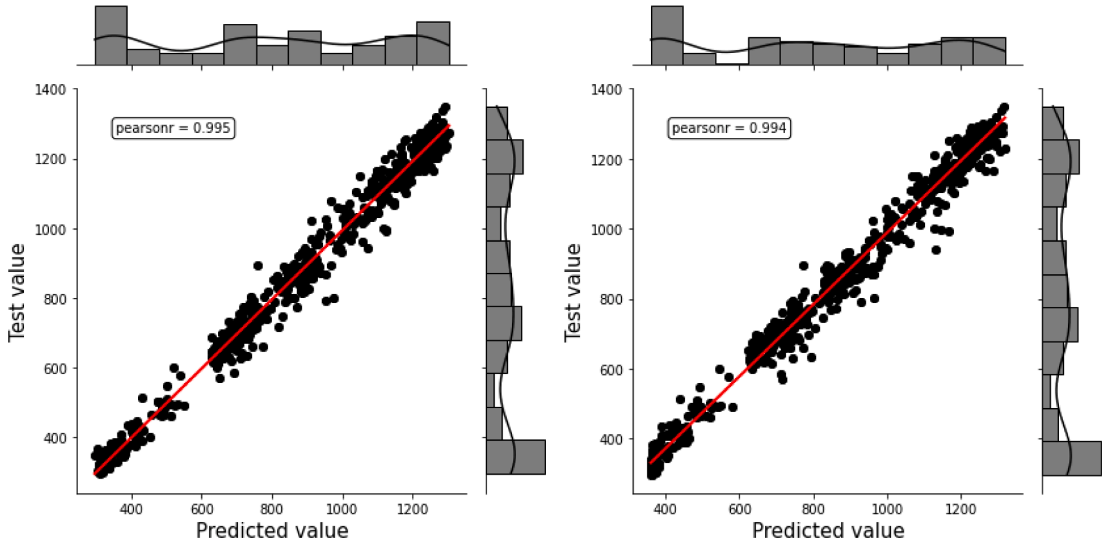


Fig. 30. Joint plot for LSTM (a) model 1 (b) model 2.

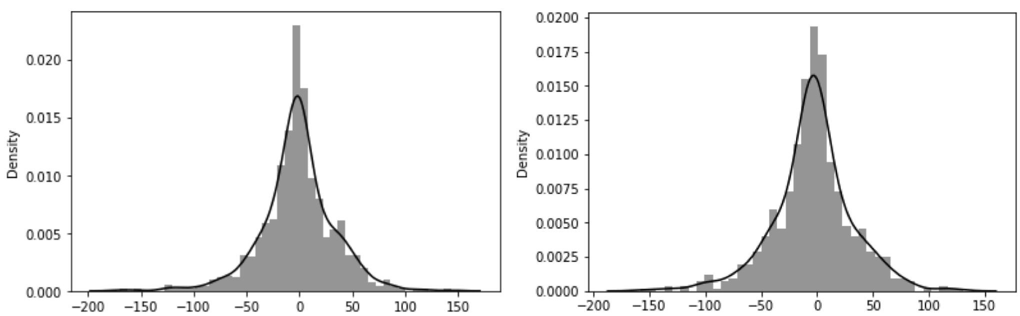


Fig. 31. Distribution plot for LSTM (a) model 1 (b) model 2.

Table 11
Results obtained from various models.

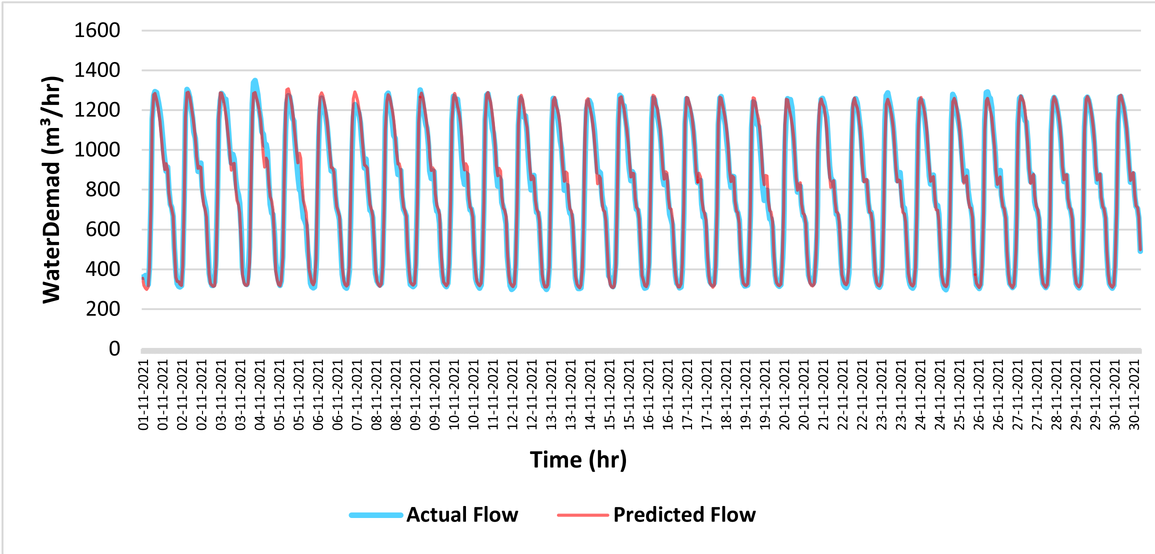
Model 1	R^2 Training data	R^2 Test data	CV score	RMSE (m^3/hr)	MSE (m^3/hr)	MAE (m^3/hr)
Linear	0.976	0.952	0.966	2.47	1.2	2.73
DT	0.970	0.963	0.952	4.53	2.1	2.86
SVR	0.964	0.973	0.951	3.91	1.5	3.31
Random Forest	0.997	0.988	0.976	2.58	1.1	1.82
XGBoost	0.981	0.989	0.976	2.52	1.1	1.79
ARIMA	0.898	0.858	0.81	4.65	2.3	5.62
KNN	0.950	0.926	0.959	3.66	0.98	2.57
ANN	0.979	0.982	0.972	0.06	0.56	0.25
LSTM	**0.992**	**0.989**	**0.981**	**0.04**	**0.1**	**0.11**

Table 12
Results obtained from various multivariate models.

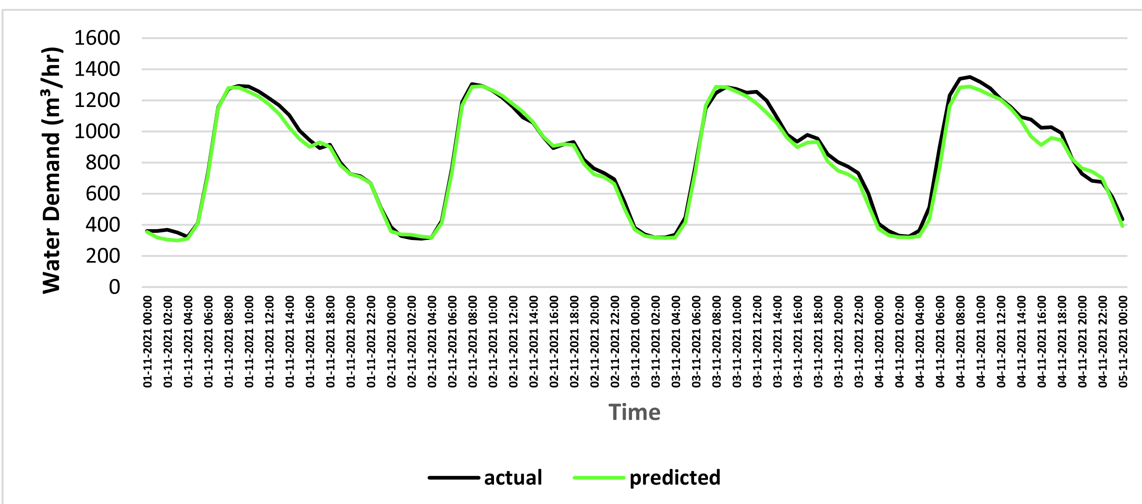
Model 2	R^2 Training data	R^2 Test data	CV score	RMSE (m^3/hr)	MSE (m^3/hr)	MAE (m^3/hr)
Linear	0.965	0.956	0.964	3.65	1.33	5.77
DT	0.96	0.937	0.9503	7.89	1.62	5.68
SVR	0.963	0.884	0.881	10.65	2.45	8.23
Random Forest	0.987	0.968	0.962	5.59	3.13	4.07
XGBoost	0.982	0.936	0.9759	7.89	1.23	5.66
ARIMA	0.832	0.791	0.811	8.82	4.82	8.03
KNN	0.979	0.895	0.957	5.64	2.85	5.82
ANN	0.978	0.981	0.965	0.193	1.93	3.48
LSTM	**0.991**	**0.988**	**0.979**	**0.14**	**1.4**	**2.96**

Machine learning and Deep learning models performed better than conventional models for water demand forecasting. The forecasting model can be developed as a univariate model when immediate short-term water demand is to be predicted to regulate water pressure, schedule pumping operations, monitor system maintenance, reduce leakage, and plan infrastructure development. Forecast based on this real consumption helps the operation and management of an existing water supply system to be efficient and effective. Multivariate model can be used for detailed analysis and forecasting for long forecast horizons when all the input parameters are available. Among all the models assessed, LSTM model was found to be the best suitable model for water demand forecasting for both the cases. The R^2 value for univariate model of LSTM was found to be 0.989 with RMSE of 0.04 m^3/hr and MAE of 0.11 m^3/hr, whereas the multivariate model of LSTM gave the better results with R^2 value of 0.988, RMSE of 0.14 m^3/hr and MAE of 2.96 m^3/hr.

With the help of this model, operation and storage efficiency can be improved in the Hubli region by predicting the future demand. The forecasted demand can be used to optimize the cost of distribution by managing the supply and storage based on the predicted demand and can help to minimize the losses in the distribution system. This short-term water demand forecast can be helpful to monitor increased water supply and storage required to fulfil future demands. It will be enhancing the capabilities of the treatment plant and water supply system to handle peak consumption as the predicted demand is available. As the future demand is available, it can be analysed, and water



(a)



(b)

Fig. 32. Comparison between actual and predicted flow using LSTM model

demand can be addressed effectively in case of short-term supply shortage or emergency demand caused by unexpected situations. Both the univariate and multivariate models produced comparable results for LSTM model and suitable model among the two can be used depending on the use and availability of the data. Thus, the LSTM forecasting model can be effectively used to forecast water demand to regulate water pressure, schedule pumping operations, monitor system maintenance, and reduce leakage in the region.

7. Future scope

The forecasting model can be effectively used to forecast water demand to regulate water pressure, schedule pumping operations, monitor system maintenance, and reduce leakage in the region. Using this short-term water demand forecasting model, the operation and management of an existing water supply system can be managed efficiently and effectively by forecasting the demand based on actual consumption and taking necessary actions as per the demand ensuring smart water management. The same model can be used for moderate and long-term forecast also to get accurate demand based on actual consumption provided the input parameters are appropriate. Furthermore, the hyperparameters for each model can be considered for further analysis as part of the future scope of the study to improve the accuracy of the results.

Funding

No specific funding has to be declared for this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could appear to have influenced the work described in this paper.

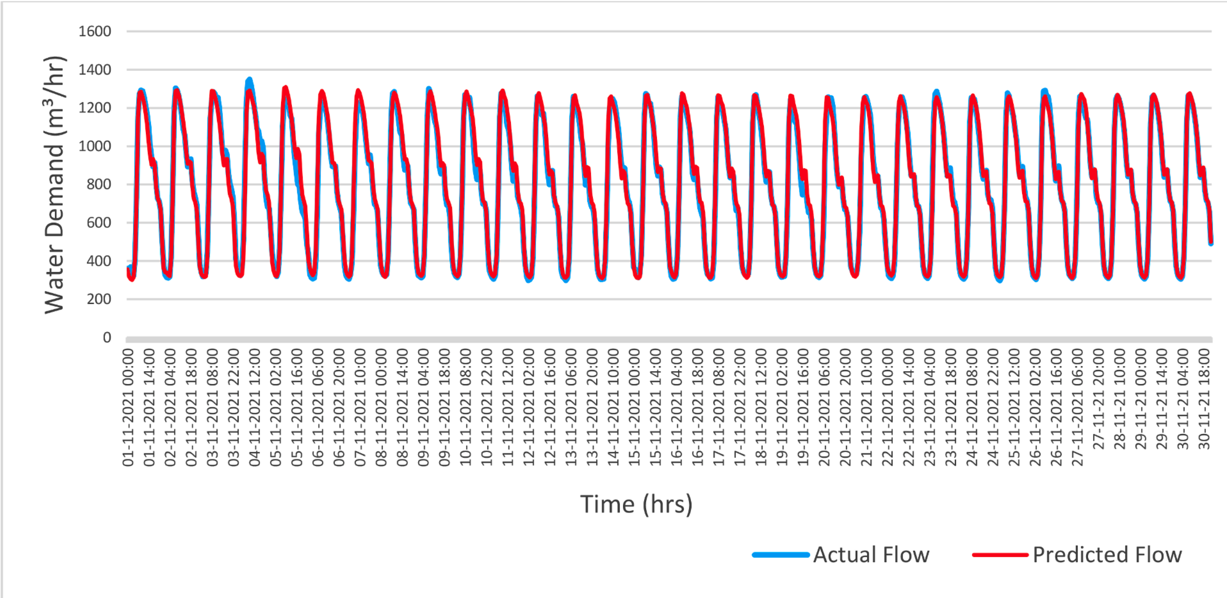


Fig. 33. Forecasted water demand.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their instructive comments, which helped improve this paper. In addition, the authors wish to thank the India Meteorological Department for providing data.

References

Abdou, A. S., Aziem, M. A., & Aboshosha, A. (2016). An efficient indoor localization system based on Affinity Propagation and Support Vector Regression. In *Sixth International Conference on Digital Information Processing and Communications (ICDIPC)*.

Adamowski, J., Chan, H. F., 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*.

Alvisi, S., Franchini, M., & Marinelli, A. (2007). A short-term, pattern-based model for water-demand forecasting. *Journal of Hydroinformatics, 9*(1), 39–50.

Aly, A. H., & Wanakule, N. (2004). Short-term forecasting for urban water consumption. *Journal of Water Resources Planning and Management, 130*(5), 405–410.

Al-Zahrani, M. A., & AbMonasar, A (2015). Urban residential water demand prediction based on artificial neural networks and time series models. *Water Resources Management, 29*, 3651–3662.

Anele, A. O., Hamam, Y., Abu-Mahfouz, A. M., & Todini, E. (2017). Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction. *MDPI, Water 2017, 9*(11), 887.

Arouna, A., & Dabbert, S. (2010). Determinants of domestic water use by rural households without access to private improved water sources in benin: A seemingly unrelated tobit approach. *Water Resources Management, 24*(7), 1381–1398.

Babel, M. S., & Shinde, V. R. (2011). Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. *Water Resources Management, 25*(6), 1653–1676.

Bakker, M., Duist, H. V., Schagen, K. V., Vreeburg, J., & Rietveld, L. (2014). Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering, 70*, 93–102.

Bakker, M., Vreeburg, J. H. G., Shagen, K. M. V., & Rietveld, L. C. (2013). A fully adaptive forecasting model for short-term drinking water demand. *Environmental Modelling & Software, 48*, 141–151.

Bárdossy, G., Halász, G., & Winter, J. (2009). Prognosis of urban water consumption using hybrid fuzzy algorithms. *Journal of Water Supply: Research and Technology e AQUA, 58*(3), 203–211.

Basani, M., Reilly, B., & Isham, J. (2004). Water demand and the welfare effects of connection: empirical evidence from Cambodia. In *In Middlebury College Working Paper Series No.0429*. Middlebury College, Department of Economics.

Bata, M., Carriveau, R., & Ting, D. S. K. (2020). Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water journal, 5*, 2.

Billings, R. B., & Jones, C. V. (2008). *Forecasting Urban Water Demand* (Second edition). American Water works Association, 2008.

Box, G., & Jenkins, G. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). San Francisco: Holden-Day: John Wiley & Sons, Inc.

Braun, M., Bernard, T., Piller, O., & Sedehizade, F. (2014). 24-hours demand forecasting based on SARIMA and support vector machines. *Procedia Eng. 2014, 89*, 926–933.

Brentan, B. M., Luvizotto, E., Jr., Herrera, M., Izquierdo, J., & Pérez-García, R. (2017). Hybrid regression model for near real-time urbanwater demand forecasting. *Joural of Computing and Applied Mathematics 2017, 309*, 532–541.

Caiado, J. (2010). Performance of combined double seasonal univariate time series models for forecasting water demand. *Journal of Hydrologic Engineering, 15*(3), 215–222.

Candelieri, A. (2017). Bicocca clustering and support vector regression for water demand forecasting and anomaly detection. *MDPI, Water 2017, 9*, 224.

Cassidy, J., Barbosa, B., Damiao, M., Ramalho, P., Ganhao, A., Santos, A., & Feliciano, J. (2020). Taking water efficiency to the next level: digital tools to reduce non-revenue water. *Journal of Hydroinformatics*.

Chang, M., & Liu, J. (2009). Water demand prediction model based on radial basis function neural network. In *Proceedings of the 2009 First International Conference on Information Science and Engineering, Nanjing, China, 26–28 December 2009* (pp. 5295–5298).

Cutore, P., Campisano, A., Kapelan, Z., Modica, C., & Savic, D (2008). Probabilistic prediction of urban water consumption using the SCEM-UA algorithm. *Urban Water Journal, 5*(2), 125–132.

Domene, E., & Saurí, D (2006). Urbanisation and water consumption: Influencing factors in the metropolitan region of Barcelona. *Urban Studies, 43*(9), 1605–1623. https://doi.org/10.1080/00420980600749969

Donkor, E., Mazzuchi, T., Soyer, R., & Roberson, J. (2012). Urban water demand forecasting: a review of methods and models. *Journal of Water Resources Planning and Management*. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314.l

Ebrahim Banihabib, Mohammad, & Mousavi-Mirkalaei, Pezhman (2019). Extended linear and non-linear auto-regressive models for forecasting the urban water consumption of a fast-growing city in an arid region. *Sustainable Cities and Society, 48*, Article 101585. https://doi.org/10.1016/j.scs.2019.101585

Fan, L., Liu, G., Wang, F., Geissen, V., & Ritsema, C. J. (2013). Factors affecting domestic water consumption in rural households upon access to improved water supply: Insights from the Wei River Basin, China. *PLoS ONE, 8*(8).

Farias, R. L., Puig, V., Rangel, H. R., & Flores, J. J. (2018). Multi-model prediction for demand forecast in water distribution networks. *MDPI*.

Fattah, J., Ezzine1, L., Aman, Z., El-Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management, 10*, 1–9.

Gagliardi, F., Alvisi, S., Kapelan, Z., & Franchini, M. (2017). A probabilistic short-termwater demand forecasting model based on the Markov Chain. *MDPI, Water 2017, 9*(7), 507. https://doi.org/10.3390/w9070507

Ghiassi, G., 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*(2), 138–146.

Gokul, P. R., Mathew, A., Bhosale, A., & Nair, A. T (2023). Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques. *Ecological Informatics, 240*, Article 102067. *pages*.

Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R., & Zhu, X. (2018). Short-term water demand forecast based on deep learning method. *ASCE 04018076-2*.

Gupta, N., Mathew, A., & Khandelwal, S (2020). Spatio-temporal impact assessment of land use /land cover (LU-LC) change on land surface temperatures over Jaipur city in India. *International Journal of Urban Sustainable Development*, 283–299.

Hussien, W. A, Memon, F. A., & Savic, D. A. (2016). Assessing and modelling the influence of household characteristics on per capita water consumption. *Water Resources Management, 30*(9), 2931–2955.

Ibrahim, T., Omar, Y., & Maghraby, F. A. (2020). *Water Demand Forecasting Using Machine Learning and Time Series Algorithms* (pp. 12–14). Pune, India: AISSMS Institute of Information Technology. Mar.

Jain, A., Varshney, A. K., & Joshi, U. C. (2001). Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resources Management, 15*(5), 299–321.

Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. *Atten Percept Psychophys, 81*, 1189–1196. https://doi.org/10.3758/s13414-019-01726-3

Kavyashree, K., & Raj, K (2020). Factors influencing urban residential water consumption in Bengaluru. In *ISEC Working Paper No. 502*.

Kenny, D. S., Goemans, C., Klein, R., Lowrey, J., & Reidy, K. (2008). Residential water demand management: Lessons from Aurora, Colorado. *Journal of the American Water Resources Association, 44*(1).

Kim, J., Lee, H., Lee, M., Han, H., Kim, D., & Kim, H. S (2022). Development of a deep learning-based prediction model for water consumption at the household level. *MDPI Water 2022, 14*, 1512. https://doi.org/10.3390/w14091512

Kofinas, D., Mellios, N., Papageorgiou, E., & Laspidou, C. (2014). Urban water demand forecasting for the island of Skiathos. *Procedia Engineering 2014, 89*, 1023–1030.

Koo, K. M., Han, K. H., Jun, K. S., Lee, G., Kim, J. S., & Yum, K. T. (2021). Performance assessment for short-term water demand forecasting models on distinctive water uses in Korea. *MDPI, Sustainability, 13*, 6056. https://doi.org/10.3390/su13116056

Maidment, D. R., Miaou, S. P., & Crawford, M. M. (1985). Transfer function models of daily urban water use. *Water Resource Research, 21*(4), 425–432.

Mathew, A., Sarwesh, P., & Khandelwal, S (2022). Investigating the contrast diurnal relationship of land surface temperatures with various surface parameters represent vegetation, soil, water, and urbanization over Ahmedabad city in India. *Energy Nexus, 5*, Article 100044. *pages*.

Mouatadid, S., & Adamowski, J (2017). Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal, 2017, 14*, 630–638.

Niu, Wen-jing, & Feng, Zhong-kai (2021). Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society*. , Article 102562. https://doi.org/10.1016/j.scs.2020.102562

Omar, S Abu Rizaiza (1991). Residential water usage: A case study of the major cities of western region of Saudi Arabia. *Water Resources Research, 27*(5).

Pu, Z., Yan, J., Chen, L., Zhirong, L., Wenchong, T., Tao, T., & Kunlun, X. (2023). A hybrid Wavelet-CNN-LSTM deep learning model for short-term urban water demand forecasting. *Frontier of Environmental Science and Engineering, 17*(2023), 22. https://doi.org/10.1007/s11783-023-1622-3

Raju, L., Gandhimathi, R., Mathew, A., & Ramesh, S. T (2022). Spatio-temporal modelling of particulate matter concentrations using satellite derived aerosol optical depth over coastal region of Chennai in India. *Ecological Informatics, 69*, Article 101681.

Rangel, H. R., Puig, V., Farias, R. L., & Flores, J. J. (2017). Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks. *Journal of Hydroinform. 2017, 19*, 1–16.

Rasifaghihi, N., Li, S. S., & Haghighat, F (2020). Forecast of urban water consumption under the impact of climate change. *Sustainable Cities and Society, 52*, Article 101848. https://doi.org/10.1016/j.scs.2019.101848

Rietveld, P., Rouwendal, J., & Zwart, B. (2000). Block rate pricing of water in Indonesia: An analysis of welfare effects. *Bulletin of Indonesian Economic Studies, 36*(3), 73–92.

Sarwesh, P., & Mathew, A (2022). Cross layer design with weighted sum approach for extending device sustainability in smart cities. *Sustainable Cities and Society, 77*, Article 103478.

Shekar, P. R., & Mathew, A (2022). Morphometric analysis for prioritizing sub-watersheds of Murredu River basin, Telangana State, India, using a geographical information system. *Journal of Engineering and Applied Science, 69*(1).

Shuang, Q., & Zhao, R. T. (2020). Water demand prediction using machine learning methods: A case study of the Beijing–Tianjin–Hebei region in China. *MDPI*, 13–310.

Sonowal, R. (2017). Factors affecting the rate of demand of water | water engineering. *Engineering notes.com*.

Strand, J., & Walker, I. (2005). Water markets and demand in central american cities. *Environment and Development Economics, 10*(3), 313–335.

Tiwari, M., Adamowski, J., & Adamowski, K (2016). Water demand forecasting using extreme learning machines. *ITP, Journal of Water and Land development, 28*(1). https://doi.org/10.1515/jwld-2016-0004

Tiwari, M. K., & Adamowski, J (2013). Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resource. Research, 49*, 6486–6507.

Vangala, S., Vadlamani, R., & Yelleti, V. (2023). ATM cash demand forecasting in an Indian bank with chaos and hybrid deep learning networks. *Expert Systems with Applications, 211*(January 2023), Article 118645. https://doi.org/10.1016/j.eswa.2022.118645. *ISSN 0957-4174*.

Villar-Navascués, R. A., & Pérez-Morales, A. (2018). Factors affecting domestic water consumption on the spanish mediterranean coastline. *The Professional Geographer, 70*(3), 513–525.

Whittington, D. (2002). Household demand for improved piped water services: Evidence from Kathmandu, Nepal. *Water Policy, 4*(6), 531–556.

Wong, J. S., Zhang, Q., & Chen, Y. D. (2010). Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. *Water Resource Research 2010, 46*, W03506.

Zhang, G., Hu, Y., Yang, D., Ma, L., Zhang, M., & Liu, X. (2022). Short-term bathwater demand forecasting for shared shower rooms in smart campuses using machine learning methods. *MDPI Water 2022, 14*(8), 1291. https://doi.org/10.3390/w14081291

Zhou, S. L., McMahon, T. A., Walton, A., & Lewis, J (2002). Forecasting operational demand for an urban water supply zone. *Journal of Hydrology, 259*(1-4), 189–202.