

Univariate water consumption time series prediction using deep learning in neural network (DLNN)

Norzanah Md Said^{1*}, Zalkan Mohd Zin¹, Mohd Nazri Ismail² and Termizi Abu Bakar³

Universiti Kuala Lumpur Malaysia France Institute, 43650 Selangor, Malaysia¹

Universiti Pertahanan Nasional Malaysia, Kem Sungai Besi, Kuala Lumpur, Malaysia²

Rai Utility Sdn. Bhd, Petaling Jaya, Selangor, Malaysia³

Received: 19-September-2020; Revised: 14-March-2021; Accepted: 16-March-2021

©2021 Norzanah Md Said et al. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The ability to predict water consumption benefits in terms of future temporal water patterns, operational reasons in distribution and decision making related to the development of water supply systems. Although, there have been numerous research work on time series forecasting, the methodology may still need further development in the context of non-linear modeling and analysis. There is many unpredictability that results in complex temporal dependence with traditional time series approaches. But now, artificial neural networks are available to help resolve these issues can improve model prediction. Therefore, this paper presents an experimental study to perform univariate water consumption for time series prediction processes using Deep Learning in Neural Network (DLNN). DLNN has abilities to learn the algorithm from previous data. It also able to process an explosion of data with a nonlinear approach, which the univariate water consumption time series prediction problem was modelled into artificial neural networks. The vast majority of learning algorithms along with DLNN are used in this study called Multilayer Perceptron (DLNN-MLP), Convolutional Neural Networks (DLNN-CNN) and Long Short-Term Memory (DLNN-LSTM) for the model prediction to generate prediction with the lowest error. Dataset from the residential water usage in SIBU Division was chosen for the experiments of this study. The performance of the DLNNs' models is measured using RMSE and state-of-the-art algorithms. The value of RMSE for a problem relates to the calculation of the lowest error. Based on the experiments result indicates that the DLNN-LSTM (RMSE:0.051) can make decent predictions for water consumption time series despite being inferior to SARIMA (RMSE:0.183). Hence, this finding implies that DLNN-LSTM can be implemented on a univariate water consumption time series prediction problem given that the problem is modelled as supervised learning.

Keywords

Water consumption, Deep neural network, Deep learning, Model prediction.

1.Introduction

Water consumption is the amount of water used by domestic and industrial water users that meet different needs of population, industry and so on. With the rapid development of industrial, population growth, particularly in urban areas, water consumption has become more and more serious with the deficiency of adequate water resources such as water shortage, water crisis or water scarcity [1–5]. Therefore, understanding water issues and patterns are of great importance for the prevention and regulation of water supply management practices [6], optimal water resources management towards sustainable growth and development [7].

Besides, it can contribute in decision making related to water supply-system development and the establishment of water rates as well. Hence, many studies have come up with forecasting techniques and models to produce a great accuracy prediction of water consumption. As noted by Altunkaynak et al. [8], an accurate prediction is required as an input for improving availability of water supply and consumer satisfaction to help water managers understand potential water usage temporal trends. In addition to Bibri [9], it also can assists the water utilities for operational reasons in distribution networks.

Due to its complexities in temporal dependencies and high dimensionalities, an accurate time series forecasting is quite challenging to be achieved. In the literature, the model prediction for water

*Author for correspondence

consumption time series has been investigated and developed based on different periodic behavior and methods [10–14]. On the other perspectives, various models have been well explored as well, such as fuzzy mathematics [15], stochastic mathematics [16–19], artificial neural networks (ANN) [20–24]. ANNs have become a popular method used in order to handle uncertain and non-linear situations. ANNs also capable of given promising a good performance for time-series prediction and classification problems. The development of IoT technologies in the field of the water domain has leads to produce an enormous amount of data such as smart meters, called Advanced Metering Infrastructure (AMI) [25–27]. The drawbacks of the traditional machine learning algorithms, the modelling complexities, the limits to prediction and computational cost are required.

Recently, deep learning (DL) is an active research that has plenty of applications in the world of statistical analysis, such as Computer Vision, Speech Recognition, Natural Language Processing. Its applicability has been examined in the areas of time series environment such as electric power system [28], medical [29] and chemical reaction [30]. However, limited studies have applied deep learning in neural network (DLNN) models to the prediction of residential water consumption (RWC).

Thus, this paper presents on the univariate water consumption (UWC) for time series prediction

processes by applying three types of Deep Learning in Neural Network (DLNN) to predict the next month of water consumption based on one-step-ahead forecasting. The main objective is to investigate and determine the UWC prediction model based on DLNN approaches. The experiments were executed on the SIBU Division dataset (with 33 monthly) and the cumulative values of accuracy were measured with the Root Mean Square Error (RMSE) and Seasonal Autoregressive Integrated Moving Average (SARIMA) as a state-of-the-art (benchmark).

2.Literature review

An overview of Artificial Neural Network (ANN) architectures that consists of multiple hidden layers ANN known as Deep Neural Network (DNN) is described. Architecture DNN has a structure deep in the temporal dimension. It consists of three layers: an input layer, an output layer and a number of hidden layers [31, 32]. According to *Figure 1*, each layer contains nodes designed to simulate neurons in a brain. Each node in each layer is fully connected to nodes in the next subsequent layer by directed edges, except for the output layer. Every node in the current layer connects and only connects to all the nodes in the adjacent subsequent layer. Moreover, there is a weight assigned to each directed edge in order to calculate output results.

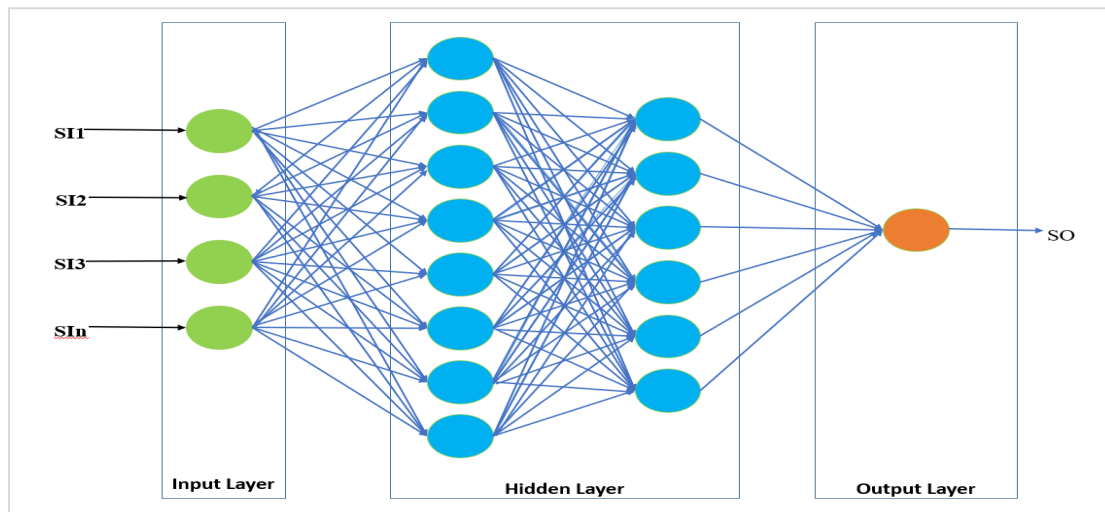


Figure 1 The structure of DNN Model

The first type of layer is called an input layer. Nodes in the input layer represent the input signals received by the data points, representing by S_{ix} , $x = \{1,$

$2, \dots, n\}$. The input values are based on the corresponding features selected from the sequence of observations obtained through repeated recorded the

amounts of water consumption over time. The second type of layer is called a hidden layer. In this case, two hidden layers are designed, which fall under the category of multiple hidden layers (DNN). For each node in a hidden layer, the node first calculates a weighted sum defined as:

$$SO=wSI+b \quad (1)$$

where w is a vector that represents weights of all directed edges incoming to a node, SI is a vector of the corresponding input values that are fed into the node, and b is a bias term. The weighted sum is passed to an activation function, and an output value is calculated. The ReLU (Rectified Linear Unit) as an activation function is applied in this article. This non-linear activation function is the most used in deep learning and popular as well.

$$f(SO)=\max\{0,SO\} \quad (2)$$

Finally, the value calculated by the activation function is utilized as the inputs of the immediately subsequent layer. The last type of layer is called as an output layer. Only one output node is required in predicting monthly UWC is shown in *Figure 1*. The output calculation in this layer is the same as the one defined in hidden layers in Equation (1) and (2). Based on the activation function, the ReLU is half rectified (from bottom). Therefore, the output, $f(SO)$ is zero when SO is less than zero and $f(SO)$ is equal to SO when SO is above or equal to zero. Bear in mind that, the output value is not restricted from 0 to 1, the activation function is which means that the weighted sum is the output.

In practice, all deep learning algorithms are neural networks, that have certain common fundamental characteristics. They are all interconnected neurons distributed in layers, only distinguished by the network architecture and sometimes the way it's trained. The vast majority of algorithms in use today, briefly explained in this article; Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM).

A Multilayer Perceptron (MLP) has the same structure as DNN, which consists of input layer, multi-hidden layer and one output layer. The input signals in the input layer are processed in forward via the hidden layer. A non-linear activation function, ReLU, transmits the summation of weighted input signals. Then, the response of network is compared with the actual observation results and the network error is calculated.

A Convolutional Neural Networks (CNN) involves the network's feedforward structure and training protocol and consists of a forward pass. A CNN is initially trained by using direct input-output weight connections. Besides, the CNN has multiple hidden layers called convolution layer, ReLU layer, pooling layer and fully connected layer. The first three hidden layers are to perform feature extraction from the data point. Meanwhile, the last layer is used to compare the response of network and calculate the network error. A Long Short Term Memory (LSTM) is a special Recurrent Neural Networks (RNNs) suitable for learning long-term dependencies [32]. It is effective at managing vanishing gradients by utilising its memory cells. *Figure 2* shows an LSTM cell with an input gate, an output gate, a forget gate and a self-recurrent neuron. Basically, the gates function is to control the interactions between adjacent memory cells and the memory cell itself. The input gate is to allow or not the input signal can alter the state of the memory cell. The output gate is used to control the state of nearby memory cells, and its connectivity to memory cell. Furthermore, the function of the forget gate is to choose to remember or forget the recently used state. As a result, the three gates' all output values are in the range of [0,1].

For the function of the memory cell in LSTM, can be expressed by:

1. Xt is the input vector to the memory cell at time t
2. Wxf, Wxi, Wxo are weight matrices
3. bf, bi, bo are bias vectors
4. ht is the value of the memory cell at time t .
 $ht - 1$ is the hidden state of the previous timestep
5. It is the values of the input gate at time t
 $It = \sigma(XtWxi + Ht - 1Whi + bi)$ (3)
6. Ft is the values of forget gate at time t
 $Ft = \sigma(XtWxf + Ht - 1Whf + bf)$ (4)
7. Ot is the values of output gate at time t
 $Ot = \sigma(XtWxo + Ht - 1Who + bo)$ (5)

A body of work exists describing the use of multiple-layer architectures to extract inherent features in data from the lowest level to the highest level. Besides, massive amounts of structure in the data can be discovered as well [33]. Therefore, as a water consumption prediction process is complicated, deep learning in the neural network can represent a model prediction of UWC time series.

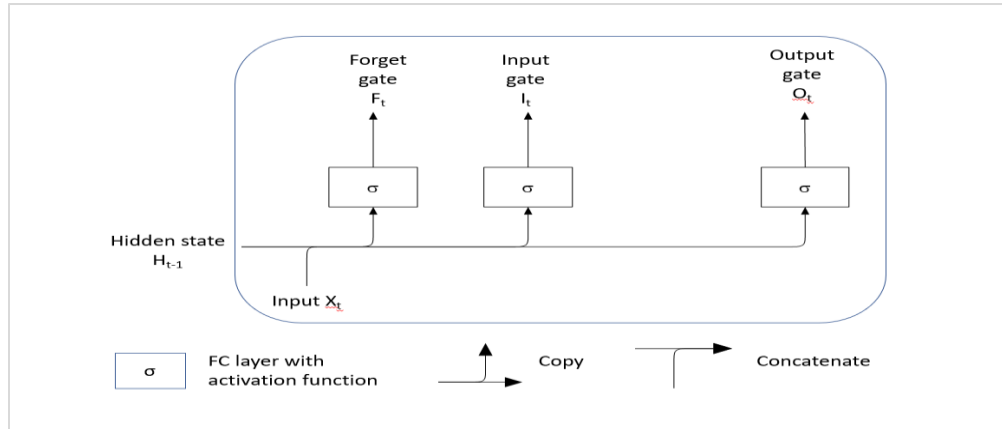


Figure 2 Data flow of unit gates in a memory cell of LSTM

3.Methods

3.1 Data collection and analysis

In this study, historical data based on the SIBU Division in Sarawak is used as a case study. The historical data was collected from the website of the official portal of Sarawak Data (<https://data.sarawak.gov.my>). The dataset consists of two variables which are labelled as Date for monthly date of water use is recorded and Consumption for the volume of water consumption per month (cubic per meter). The dataset also consists of 33 observations for each record on the monthly domestic water consumption data from January 2017 until September 2019. The sample of the raw dataset for SIBU Division domestic water consumption is shown in *Table 1*.

According to the *Table 2* shows that this study focuses on univariate time series. The dataset comprised of a single series of observations that attaches a time period to each value. An important component of the study is an observed time series which is the values of consumption. Then, the approach can be seen by applying two datasets were fed into the algorithm, called non-stationary (NS) and stationary (S) time-series data. The NS can be

described by using raw data without pre-processing process and the characteristics of time series have trend and seasonality. While S can be explained by require pre-processing and make the stationary process of time series to eliminate the trend and seasonality.

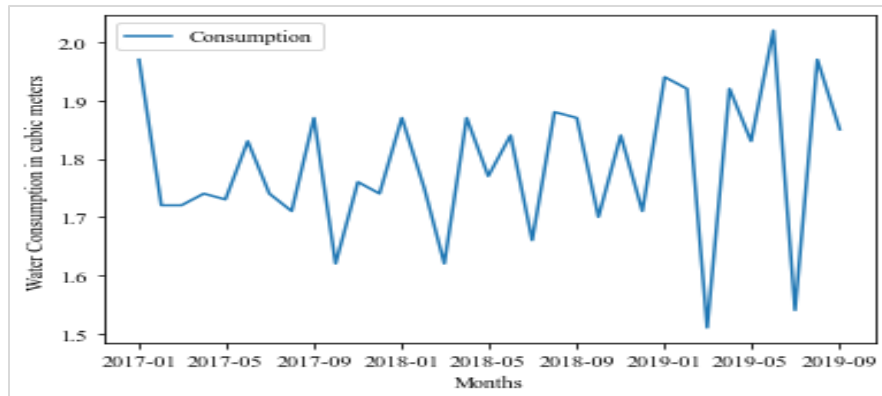
Figure 3 shows the observed monthly water consumption data using NS time series data in SIBU Division. According to the plot, its appearance a slight increase in trend and indicated periodical behaviour. Based on this visual inference describes that its statistical properties such as mean, variance change over interval time. The high consumption of waters in January, Mei and September; and the low consumption of waters in Mac, July and October are determined. Therefore, it can be assumed to NS time series data. In the viewpoint of the statistical tests named Augmented Dickey-Fuller (ADF) test, which ADF test more (-0.392) than critical values (1% - -3.78, 5% --3.01, 10% --264) and P-Value (0.911305) is more than 0.05, indicates the NS[34]. Thus, the dataset of the SIBU Division is known as NS, is shown in *Figure 4*.

Table 1 Sample of raw dataset, SIBU Divison water consumption

Date	Consumption
2017-01	1.97
2017-02	1.72
2017-03	1.72
2017-04	1.74
2017-05	1.73
2017-06	1.83
2017-07	1.74
2017-08	1.71
2017-09	1.87

Table 2 Sample of water consumption time series data

Date	Consumption
2017-01-01	1.97
2017-02-01	1.72
2017-03-01	1.72
2017-04-01	1.74
2017-05-01	1.73
2017-06-01	1.83
2017-07-01	1.74
2017-08-01	1.71
2017-09-01	1.87

**Figure 3** Visualizing monthly water consumption in SIBU division

```

ADF Statistic: -0.392731
p-value: 0.911305
Critical Values:
1%: -3.7701
5%: -3.005
10%: -2.643

```

Figure 4 The statistical test (ADF Test) for the first dataset, raw data

```

ADF Statistic: -4.406559
p-value: 0.000289
Critical Values:
1%: -3.809
5%: -3.022
10%: -2.651

```

Figure 5 The statistical test (ADF Test) for the second dataset, preprocessed data

Then, the second dataset is created using a differencing technique to make the stationary process that known as stationary (S) time-series data. According to *Figure 5* shows that the ADF test shows less (-4.407) than all critical values and P-Value (0.000289) is less than 0.05.

The next step is each dataset was split into training and testing dataset using 70% and 30% of the total data records. The models are tested by testing on new sets of data, which have not run through the training process in order to lead more reliable evaluation. The

dataset has been examined without missing values using approaches by Said et al. [35].

3.2 Model development

A model is required to learn from the series of past observations of water consumption to predict the observation at the next time step. This kind of approach is well known as one-step ahead forecasting. In predicting UWC time series, three types of model based on DLNN; MLP, CNN and LSTM methods are proposed is shown in *Table 3*. On the other hand, these models are implemented due to provide adequately accurate and reliable performance. Thus, the application of these models extensively embedded with the feature selection using a machine learning algorithm to seek feature selection on time series input variables. In this study, the feature selection of lag observation using the Random Forest (RF) algorithm based on wrapper-based feature selections (recursive feature elimination (RFE)) is adopted [36]. The feature ranking is applied to design the WCP architecture using DLNN approaches. This is an important step in the creation of an effective model prediction [37–39]. Then, the model of feature selection (MoFS) as input variables for predicting UWC is proposed as shown in *Table 4*. The system architecture of the water consumption prediction (WCP) model is constructed on the basis

of Equation (6). Then, the WCP models are created is shown in *Table 5*. The models are developed based on adaptation by Firat et al. [37].

$$SO = f(SO(IM1), (SO(IM2), \dots (SO(IMn)) \quad (6)$$

Table 3 Models developed in the experiments with short description and abbreviation used

Model	Description	Abbr Used
Multilayer Perceptron	MLP with non-stationary	DLNN-MLP-NS
Multilayer Perceptron	MLP with stationary	DLNN-MLP-S
Convolution Neural Network	CNN with non-stationary	DLNN-CNN-NS
Convolution Neural Network	CNN with non-stationary	DLNN-CNN-S
Long Short Term Memory	LSTM with non-stationary	DLNN-LSTM-NS
Long Short Term Memory	LSTM with stationary	DLNN-LSTM-S
Seasonal Autoregressive Integrated Moving Average (SARIMA)	The state-of-the-art on traditional time series forecasting	SARIMA

Table 4 The proposed model of feature selection (MoFS) based on random forest algorithm

MoFs	Input
IM1	SO (-t1)
IM2	SO (-t1), SO (-t2), SO (-t3), SO (-t4), SO (-t5)
IM3	SO (-t1), SO (-t2), SO (-t3), SO (-t4), SO (-t5), SO(-t6)
IM4	SO (-t1), SO (-t2), SO (-t3), SO (-t4), SO (-t5), SO(-t6), SO(-t7)

Table 5 WCP model structures

Input variables	Model structures
MoFs	DLNN-MLP-NS
MoFs	DLNN-MLP-S
MoFs	DLNN-CNN-NS
MoFs	DLNN-CNN-S
MoFs	DLNN-LSTM-NS
MoFs	DLNN-LSTM-S

3.3 Performance criteria

The statistical criteria, Root Mean Square Error (RMSE) is calculated to compare and evaluate the results of the DLNN-MLP, DLNN-CNN and DLNN-LSTM models. These models compete to return the lowest error and then, a skilful or good performance can be determined. The Equation for RMSE is shown in Equation (7).

The values of RMSE for the prediction is obtained through the square root of the average error calculated. The RMSE statistic indicates a model's ability to predict a value away from the actual value. Besides that, the model evaluation is implemented the repeat evaluation of each model configuration to ten (10) times and report the error as the average error across each assessment. In addition to using metric, the performance of DLNN-MLP, DLNN-CNN and DLNN-LSTM models are also evaluated by comparing it with a benchmark model. This paper used the Seasonal AutoRegressive Integrated Moving Average (SARIMA) as state-of-the-art. SARIMA is a popular and widely used statistical model for seasonal time series prediction. It contains six

parameters; three parameters as non-seasonal called p, d, and q and the other three parameters called seasonal (P, D, Q). The SARIMA setting that is used in this paper is SARIMA [(1,1,1), (1,1,1,12), 'c']. In the experiment, the SARIMA model obtained 0.183 in terms of RMSE values. A model is a skillful or better performance other than SARIMA or other models when returning the lowest error of RMSE.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\widehat{SO}_t - SO_t)^2}{n}} \quad (7)$$

where;

\widehat{SO}_t : is the predicted consumption values at time, t

SO_t : is the observed consumption values at time, t

t : is the number every time point from $t = 1$ to $t = n$

n : is the number of consumption observations

3.4 Experimental setup

In this study, Python 3 programming language using Jupyter Notebook IDE 6.0 for WCP model was developed. The experiments were executed on a notebook with properties such as Intel® Core™ i7-3517U CPU @ 1.9GHz processor, 8.00GB RAM, Windows 10 Home Single Language 64-bit operating

system, NVIDIA GeForce 635M graphic card, and 1TB HDD storage. The WCP models were trained and tested on a non-stationary and stationary time series data analyzed.

4.Results

4.1Data preprocessing results

In this study, RFE technique with RF algorithm is used to select UWC feature selection which obtains

the score of each feature and select those features with the smallest score. Based on *Figure 6*, four selected features are determined. The lag observation of 1 (-t1), lag observation of 5 (-t5), lag observation of 6 (-t6), and lag observation of 7 (-t7) are used to construct the WCP models; DLNN-MLP, DLNN-CNN and DLNN-LSTM.

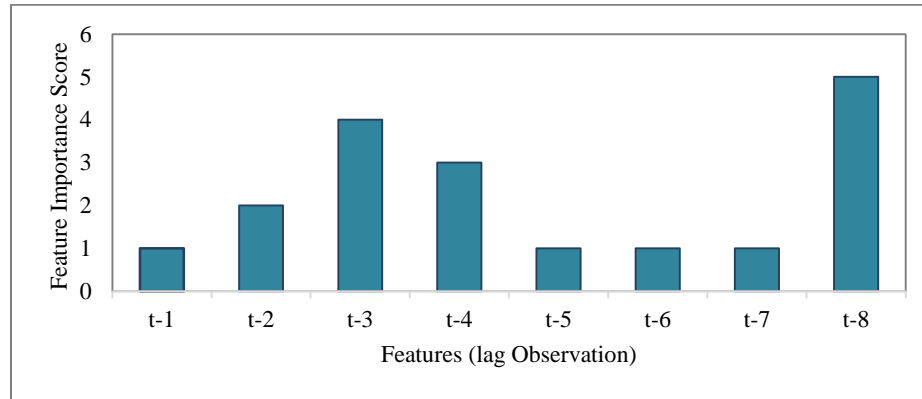


Figure 6 The feature selection rank for each of the eight lag observations

4.2 UWC prediction using DLNN

The performances of the models (*Table 5*) are shown in *Table 6*. Here the state-of-the-art algorithms using SARIMA models obtained 0.183 in term of its RMSE. In the viewpoint of MoFS for NS, the results indicate that the performances of IM2 model for DLNN-MLP, DLNN-CNN and DLNN-LSTM are the lowest score in term of the prediction accuracy compare with others, 0.181, 0.188 and 0.187. Based on this finding, it can be described that DLNN-MLP is a better performance in term of the lowest prediction error score. On the other hand, MoFS for S indicates that IM3 for DLNN-MLP, and IM2 for DLNN-CNN and DLNN-LSTM demonstrated to

have good performances in term of their prediction accuracy; 0.116,0.076 and 0.051. This shows that the prediction accuracy of MoFS of each dataset will return to the IM2 which indicates input variables of lag observation of 5, (-t5) is selected. Then, the lowest score error, IM2 for each dataset in *Figure 6(a)* and *6(b)* were compared. Results show that IM2 DLNN-LSTM performs better in *Figure 6(b)* in term of RMSE (0.051) with respect to other models, IM2 DLNN-MLP (0.181) in *Figure 6(a)*. This approach outperforms other methods in term of RMSE value obtained.

Table 6 Performances measures of water consumption prediction based on three different of DLNN models

a) Using raw data : NS

MoFS	DLNN-MLP	DLNN-CNN	DLNN-LSTM
IM1	0.289	0.286	0.256
IM2	0.181	0.188	0.187
IM3	0.188	0.193	0.196
IM4	0.183	0.190	0.195

b) Using preprocessed data: S

MoFS	DLNN-MLP	DLNN-CNN	DLNN-LSTM
IM1	0.151	0.103	0.103
IM2	0.132	0.076	0.051
IM3	0.116	0.104	0.056
IM4	0.117	0.114	0.088

Furthermore, the performance of the error can be expanded using a box and whisker plot are shown in *Figure 7* and *Figure 8*. The statistical graph method is used to explain the spread of errors for visualizing the error distribution. The findings show that the DLNN-MLP (7a), DLNN-CNN (8b) and DLNN-LSTM (7c) are indicated as negative skewed distributed. The negative skewed distributed means that the spread is biased towards lower error values. Other findings show that the DLNN-CNN (7b) and DLNN-MLP (8a) are indicated as positive skewed

distributed. The positive skewed distributed is described as the spread biased towards larger error values. Finally, only the DLNN-LSTM (8c) is indicated as normally distributed. The result from this model is likely to be better than those of the other models. Due to that, the DLNN-LSTM using stationary data is considered better model performances of UWC in the short term than those models.

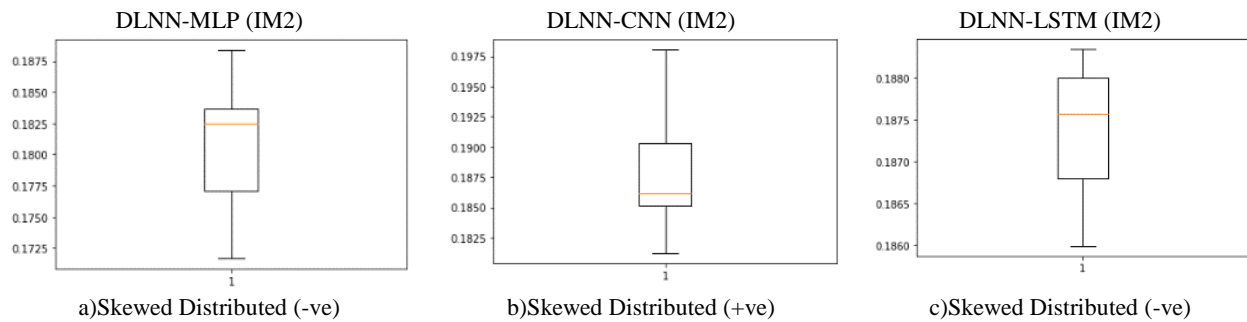


Figure 7 A spread of the performance for three different of DLNN Models using NS based on skilful MoFS using a box and whisker plot

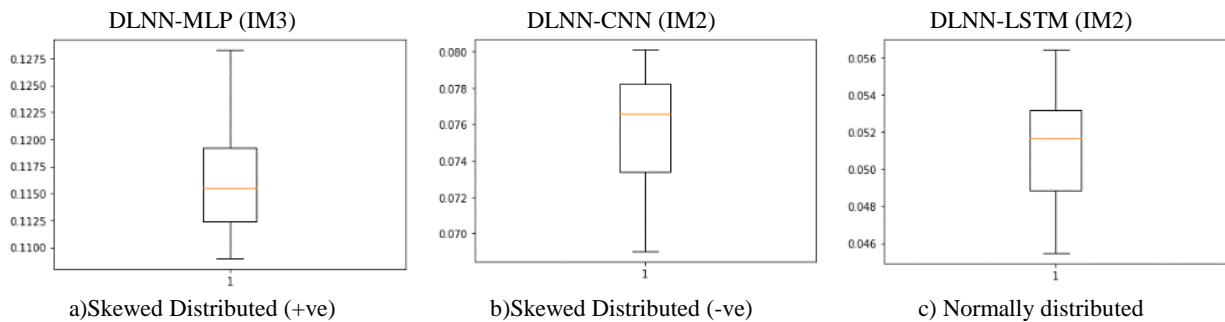


Figure 8 A spread of the performance for three different of DLNN Models using S based on skilful MoFS using a box and whisker plot

5. Discussion

The method presented in this paper is sufficiently general to be applied in the forecast of the univariate time series even with data limitations. This finding can also occur in the case of direct use of raw data when the DLNN approaches is applied. However, if accurate prediction accuracy is emphasized, a stationary process approach is the best choice of action. Based on the case study, the UWC time series forecast using the stationary process approach provided significant results that obtain the lowest forecast error score. Other than that, this was originally designed for a water utility domain but could be used in any business operations and management as well.

5.1 Limitation of study

Despite the success demonstrated, significant limitation to be considered are identified such as feature selection technique and DNN model structure used. In fact, the right choice of feature selection algorithm significantly given an effect to the model performance. In this paper, the rfe algorithm is used and obtained reasonable results. However, using different feature selection algorithm, will obtained different findings in term of model performance. To the best of our knowledge, rfe technique is the most popular used for feature ranking either in regression or classification problem. In the perspective of DNN model structure designed, this paper used two hidden layers. Determining more than two hidden layers will benefited for the layers to learn complex

representation which means for automatic feature engineering. In term of simplicity and limitation of size sample, this study preferred to apply two hidden layers. Bear in mind that, applying two hidden layers be able to represent an arbitrary decision with rational activation functions. On the other hand, it can approximate any smooth mapping to any prediction accuracy.

6. Conclusion and future work

In summary, this paper presents the abilities of Deep Learning Neural Network (DLNN) in understanding past patterns of water consumption observations to predict future water consumption behavior in the SIBU division using one-step-ahead forecasting. The DLNN techniques considered include Multilayer Perceptron (MLP), Convolution Neural Network (CNN) and Long Short Term Memory (LSTM). For this purpose, six models of DLNN techniques based on four models of features selection as input variables in predicting UWC values are constructed. The MoFS comparison showed that the performance of the IM2 model which consists of lag observation of 5 (-t5) have demonstrated a good performance compare other models investigated with respects to the RMSE value. On the other hand, the comparison of predictive models' results in DLNN-MLP, DLNN-CNN and DLNN-LSTM models in NS and S time series data, the performance of the IM2 DLNN-LSTM model in S is slightly better than those of others and the state-of-the-art as well. Based on the viewpoint of the statistical graph method, the IM2 DLNN-LSTM in S model is consistent in terms of its errors from a symmetrical distribution. Future directions are related to the improvement of model's prediction through the model-based optimization techniques will be discussed later.

Acknowledgment

The authors are grateful to the RAI Utility Sdn Bhd for providing the necessary support to carry out the work presented in this manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Cutore P, Campisano A, Kapelan Z, Modica C, Savic D. Probabilistic prediction of urban water consumption using the SCEM-UA algorithm. *Urban Water Journal*. 2008; 5(2):125-32.
- [2] Shuangli W, Mou L, Shen D, Jianjun W, Hui X. Classification calculation on water consumption of urban water supply network based on clustering analysis. In *World Automation Congress 2012* (pp. 1-4). IEEE.
- [3] Liang B. Urban annual water consumption prediction using artificial neural network. In *applied mechanics and materials 2013* (pp. 1008-11). Trans Tech Publications Ltd.
- [4] Wang ZH, Xuan CG, Fu Q. Urban water consumption forecasting based on projection pursuit. In *advanced materials research 2012* (pp. 1334-8). Trans Tech Publications Ltd.
- [5] Yang T, Shi P, Yu Z, Li Z, Wang X, Zhou X. Probabilistic modeling and uncertainty estimation of urban water consumption under an incompletely informational circumstance. *Stochastic Environmental Research and Risk Assessment*. 2016; 30(2):725-36.
- [6] Donkor EA, Mazzuchi TA, Soyer R, Alan Roberson J. Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management*. 2014; 140(2):146-59.
- [7] Altunkaynak A, Özger M, Çakmakci M. Water consumption prediction of Istanbul city by using fuzzy logic approach. *Water Resources Management*. 2005; 19(5):641-54.
- [8] Altunkaynak A, Nigussie TA. Monthly water consumption prediction using season algorithm and wavelet transform-based models. *Journal of Water Resources Planning and Management*. 2017; 143(6):04017011.
- [9] Bibri SE. The IoT for smart sustainable cities of the future: an analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society*. 2018; 38:230-53.
- [10] Benítez R, Ortiz-Caraballo C, Preciado JC, Conejero JM, Sánchez Figueroa F, Rubio-Largo A. A short-term data based water consumption prediction approach. *Energies*. 2019; 12(12):1-24.
- [11] Shah S, Hosseini M, Miled ZB, Shafer R, Berube S. A water demand prediction model for central indiana. In *proceedings of the AAAI conference on artificial intelligence 2018* (pp.7819-24).
- [12] Velasco LC, Granados AR, Ortega JM, Pagtalunan KV. Medium-term water consumption forecasting using artificial neural networks. In *conference of the science council of Asia, national research council of the Philippines 2017*:1-8.
- [13] Aly AH, Wanakule N. Short-term forecasting for urban water consumption. *Journal of Water Resources Planning and Management*. 2004; 130(5):405-10.
- [14] Candelieri A, Soldi D, Archetti F. Short-term forecasting of hourly water consumption by using automatic metering readers data. *Procedia Engineering*. 2015; 119:844-53.
- [15] Bigui W, Hongwei Z. Traversal grey model for water consumption prediction. In *international symposium on geomatics for integrated water resource management 2012* (pp. 1-4). IEEE.
- [16] Wang W, Jiang J, Fu M. An enhanced differential evolution based grey model for forecasting urban water consumption. In *proceedings of the Chinese control conference 2014* (pp. 7643-8). IEEE.

- [17] Liu BZ, Wang DW. Grey regressive prediction method of urban life water consumption. In the Chinese control and decision conference 2014 (pp. 2913-8). IEEE.
- [18] Zhou H, Huang J, Yuan Y, Tang B. Prediction of water consumption in hospitals based on a modified Grey GM (0, 1| sin) model of oscillation sequence: the example of Wuhan City. *Journal of Applied Mathematics*. 2014:1-7.
- [19] Froelich W, Magiera E. Forecasting domestic water consumption using bayesian model. In *intelligent decision technologies 2016* (pp. 337-46). Springer, Cham.
- [20] Ren Y, Xing T, Chen X, Xu E, Zhao Y. Based on process neural network learning algorithm for prediction of urban water consumption. In *international conference on computing, control and industrial engineering 2010* (pp. 34-7). IEEE.
- [21] Wang L. Improving artificial neural network models for predicting hourly water consumption (Master's thesis, University of Waterloo). 2018.
- [22] Chen G, Long T, Xiong J, Bai Y. Multiple random forests modelling for urban water consumption forecasting. *Water Resources Management*. 2017; 31(15):4715-29.
- [23] Farah E, Abdallah A, Shahrour I. Prediction of water consumption using artificial neural networks modelling. In *MATEC web of conferences 2019* (p. 01004). EDP Sciences.
- [24] Zhang W, Bai Y. Prediction of water consumption using NARX neural network based on grey relational analysis. In *international conference on sensing, diagnostics, prognostics, and control 2018* (pp. 471-5). IEEE.
- [25] Pesantez JE, Berglund EZ, Kaza N. Smart meters data for modeling and forecasting water demand at the user-level. *Environmental Modelling & Software*. 2020; 125:104633.
- [26] Tu C, He X, Shuai Z, Jiang F. Big data issues in smart grid—a review. *Renewable and Sustainable Energy Reviews*. 2017; 79:1099-107.
- [27] Xenochristou M. Water demand forecasting using machine learning on weather and smart metering data. 2019.
- [28] Wen L, Zhou K, Yang S. Load demand forecasting of residential buildings using a deep learning model. *Electric Power Systems Research*. 2020; 179:106073.
- [29] Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. *Journal of Biomedical Informatics*. 2017; 69:218-29.
- [30] Fooshee D, Mood A, Gutman E, Tavakoli M, Urban G, Liu F, et al. Deep learning for chemical reaction prediction. *Molecular Systems Design & Engineering*. 2018; 3(3):442-52.
- [31] Kubat M. *An introduction to machine learning*. Springer International Publishing AG; 2017.
- [32] Graves A. Long short-term memory. In *supervised sequence labelling with recurrent neural networks 2012* (pp. 37-45). Springer, Berlin, Heidelberg.
- [33] <https://www.packtpub.com/product/python-deep-learning-second-edition/9781789348460>. Accessed 26 October 2020.
- [34] Kwiatkowski D, Phillips PC, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*. 1992; 54(1-3):159-78.
- [35] Said NM, Zin ZM, Ismail MN, Bakar TA. Comparative analysis of missing data imputation methods for continuous variables in water consumption data. *International Journal of Advanced Trends in Computer Science and Engineering*. 2019; 8(1.6):471-8.
- [36] Tyralis H, Papacharalampous G. Variable selection in time series forecasting using random forests. *Algorithms*. 2017; 10(4):1-25.
- [37] Firat M, Turan ME, Yurdusev MA. Comparative analysis of neural network techniques for predicting water consumption time series. *Journal of Hydrology*. 2010; 384(1-2):46-51.
- [38] Koller D, Sahami M. Toward optimal feature selection. *Stanford InfoLab*; 1996:1-15.
- [39] Kumar V, Minz S. Feature selection: a literature review. *SmartCR*. 2014; 4(3):211-29.



Norzanah Md Said, received a Bachelor degree in Computer Science in 1996 from Universiti Putra Malaysia (UPM), Masters in Information Technology (Computer Science) in 2002 from Universiti Kebangsaan Malaysia (UKM). Currently, she is doing PhD in Industrial Automation at Universiti Kuala Lumpur, Malaysia. Her current research interests are Machine Learning, Artificial Intelligence and Deep Learning
Email: norzanah@unikl.edu.my



Dr Zalhan Mohd Zin is a Director UniKL International Office (UIO) Universiti Kuala Lumpur in Malaysia. He holds a PhD in Artificial Intelligence from Newcastle University (2015), MSc in Artificial Intelligence from Universiti Teknologi Malaysia, Malaysia (2007), and BSc in Computer Programming Specific Applications from the Institute Universite Professionalise (IUP) Calais, France (1999). His research interests are Artificial Intelligence, Machine Learning and Deep Learning.
Email: zalhan@unikl.edu.my



Assoc. Prof. Dr Mohd Nazri Ismail is a Deputy Dean (Academic and Undergraduate), National Defence University of Malaysia. He holds a PhD in Computer Science (Industrial Computing – Computer Network) from Universiti Kebangsaan Malaysia (2010), MSc in Engineering Science

(Communication & Network) from Universiti of Multimedia, Malaysia (2007), and BSc in Computer Science from Universiti Kebangsaan Malaysia (1995). His research interests are Digital Forensic, Network Security, IoT, WSN and MANET.

Email: m.nazri@upnm.edu.my



Mr Termizi Abu Bakar is a Managing Director at Rai Utility Sdn Bhd from Malaysia. He is a BSc holder in Computing Information System from University of Lincoln, United Kingdom (2004). He has been involving in managing related to business development, project coordination, and

providing consultancy services related to his area of expertise and interest. His current project interest is Non-Revenue Water Projects.

Email: termiziabubakar@gmail.com