

Research Article

Prediction of Fault Occurrences in Smart City Water Distribution System Using Time-Series Forecasting Algorithm

Maheswari Chenniappan ¹, **Divya Gnanavel** ², **Kavi Priya Gunasekaran** ²,
R.R. Rajalakshmi ³, **A.S Ramya** ⁴, **Albert Alexander Stonier** ⁵, **Geno Peter** ⁶,
and Vivekananda Ganji ⁷

¹Department of Mechatronics Engineering, Kongu Engineering College, Perundurai, India

²Department of Information Technology, Hindusthan College of Engineering and Technology, Coimbatore, India

³Department of Computer Science Engineering, Kongu Engineering College, Perundurai, India

⁴Department of Civil Engineering, Sanskrithi School of Engineering, Puttaparthi, India

⁵Department of Electrical and Electronics Engineering, Kongu Engineering College, Perundurai, India

⁶CRISD, School of Electrical and Technology, University of Technology Sarawak, Sibu, Malaysia

⁷Department of Electrical and Computer Engineering, Debre Tabor University, Debra Tabor, Ethiopia

Correspondence should be addressed to Vivekananda Ganji; drvivek@bhu.edu.et

Received 24 February 2022; Revised 11 April 2022; Accepted 11 June 2022; Published 5 July 2022

Academic Editor: Punit Gupta

Copyright © 2022 Maheswari Chenniappan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The proposed research work is focused on forecasting the future requirements of water supply based on the current requirement of water and also identifying the possibility of occurrences of cracks and leaks using the ARIMA (autoregressive integrated moving average) model. The experiments were conducted using real-time experimental hardware. The pressure data obtained and their p -value is less than 0.05, which represents the stability of the data in the ARIMA model. The forecasted pressure data range between 0.451379 N/m² and 2.022273 N/m². The frequency of the forecasted pressure ranges between 1.706869 N/m² and 3.065836 N/m² (maximum peak) and -0.81046 N/m² and 1.042164 N/m² (minimum peak). Forecasted data of pressure at damaged condition lie between 2.880788 N/m² and 3.29797 N/m² and frequency ranges between 4.866227 N/m² and 5.664348 N/m². Similarly, future forecasted data of water requirement for the next 1 year range between 614.6292 (liters/week) and 620.0099 (liters/week), the frequency of the forecast value with maximum ranging from 617.0086 (liters/week) to 628.5465 (liters/week), and the minimum peaks ranging from 611.0967 (liters/week) to 612.2914 (liters/week). The above data are for a single water distribution pipeline.

1. Introduction

Water pipelines face significant problems as a result of chemical leaks, fires, and deformations such as particle accumulation, corrosion, and cracks caused by a variety of factors. The above leads to serious consequences, as the distribution of clean water is one of the major objectives and the whole world depends on it. Hopkins states that the water supply framework today comprises of foundation that gathers, oversees, stores, and conveys water from water sources to shoppers. Because of the absence of new common water sources and an inexorably developing populace, inventive water assets the executive's approaches are required.

[1] Water conveyance frameworks are right now confronting various significant difficulties, including maturing foundation, the interest for consumable water, protecting consumable water quality, debased foundation because of framework failures, ecological concerns, and rising energy costs [2]. Another significant issue facing water utilities is spillage; when it is not noticed long ago, most endeavors address this issue happened after a break or hole had happened. [3, 4] Many researchers studied on the mean-shift algorithm with Gaussian's profile and made applications in the tracking system for better performance in the field of tracking objects. [5, 6] Monica et al. brief that the breaks and holes were fundamentally brought about by enormous varieties in pipe

pressure inside a water conveyance framework [7, 8]. This proposed research is related to [9–12] oil pipeline transportation monitoring using a machine learning algorithm, which is the base work to carry out the present research. What's more, subsequently, in this task, research has been led to recognize the issues that might be happening in the water circulation frameworks ahead of time utilizing time arrangement estimation calculation. The ARIMA model is used for predicting the future based on the past history of variations in datasets. However, it cannot be used to predict highly uncertain long predictions [13]. Jianbin Huang et al. clarify that the time arrangement anticipation is the way toward fitting models to chronicled information and afterward utilizing those models to foresee future perceptions. [14] In the initial step, past perceptions are accumulated and dissected to make a numerical model that catches the arrangement of hidden information-age measures. The model has been utilized to conjecture future occasions in the subsequent stage. This technique is particularly helpful when there is no sufficient logical model accessible. [15, 16] There are various kinds of time-series forecasting broadly utilized. [17] A study used an integrated genetic algorithm (GA), an algorithm which is used for forecasting monthly electrical energy consumption. [18] A detailed review study has been carried out by the authors [19] about the different forecasting models. Among them, the ANN models, support-vector machine models, and ARIMA models were mostly preferable. The present research work uses the ARIMA model for anticipating the future requirements of the water management system. This examination of the line breaks and harms is used to screen the gas or oil pipelines that regularly get harmed and dirty the climate. [20–22] Ling Yang et al. state that the anticipation should be possible by utilizing the both occasional and nonoccasional ARIMA models. [23]. Time arrangements are utilized in the assortment of fields, including numerical money, fabricating, occasion information, IoT information, and some other space of applied science and designing that require worldly estimations. [24–26] As the quickest developing fragment of the information-based industry, time arrangement DBMS will bear witness to the business' developing requirement for time arrangement determination. [27, 28]. Stormwater and drainage management system with an IoT module was developed by the researchers to predict the drainage using machine learning algorithms. The results show that a well-trained algorithm can predict the drainage situations. Hence, the proposed research work is focused on using a time series-based ARIMA model explored through the R studio software, which is one of the widely used platforms for stock predictions. The detailed mathematical study and the results are provided in the subsequent sections to ensure the prediction accuracy.

2. Methodology

2.1. Experimental Hardware for Water Distribution System. The real-time data are collected from the experimental hardware for the water distribution system, which consists of various components' process tank, reservoir tank, I/P converter, pressure control valve, gate valve, manual hand

valve drain valve, pressure gauge, flow control valve, orifice flow meter, etc., [29] as shown in Figure 1. The experimental hardware comprises three different water control loops from the reservoir to the distribution station for measuring pressure, flow, and level in the water distribution system [30]. Pressure, flow, and level are measured using sensors, and their corresponding data will be monitored using a master control engineering station [31, 32]. The control mechanism involves two-layer control, namely, operator work station and engineering station. The former one has the highest control capability than the second one.

The output of the transmitters is fed as input to the input/output hub module, which is interfaced with the field control station. The process flow diagram for the experimental setup is shown in Figure 2. The flowing water in the setup shown in Figure 2 is discussed as follows [33, 34]. The water is sucked from the water reservoir and is fed into the process level tank with the use of a pump and transmitted to the pipeline system. Different setups for level recording, flow recording, and pressure recording are made to record the real-time data from the water distribution system [35, 36] as shown in Figures 3, 4, and 5. The water distribution system is pilot scaled and used for real-time data monitoring based on pressure and flow sensors' feedback.

The data are obtained in the form of graphs and datasets by using the front end of HMI [37] as shown in Figures 4 and 6. The water is again drained into the water reservoir after the recordings are made, and the total system is shut down. Auto-tuning PID controller is used in the experimental hardware, and it is used for flow and pressure control [38, 39].

2.1.1. The ARIMA Model for the Future Prediction. The autoregressive integrated moving average (ARIMA) model is used for future predicted time-series values based on current pressure and flow data in the water distribution system. The ARMA model and ARIMA models have many similarities. However, the general autoregressive model estimates predictions using previous values of the dependent variable [40, 41]. The ARIMA uses the linear regression model, which uses lags as predictors. The ARIMA model uses a forecasting equation based on the time-series regression equation, and it is represented as follows:

Predicted value M = weighted sum of recent values of M + weighted sum of recent values of the errors.

The time-series forecasting functional equation is constructed as follows [33, 34]. First, let the predicted value M indicates n , which is the n th difference of M , and it can be represented as follows:

$$\begin{aligned} \text{When, } n = 0: M_t &= M_t, \\ \text{When, } n = 1: M_t &= M_t - M_{t-1}. \end{aligned} \quad (1)$$

When $n = 2$,

$$\begin{aligned} M_t &= (M_t - M_{t-1}) - (M_{t-1} - M_{t-2}) \\ &= M_t - 2M_{t-1} + M_{t-2}. \end{aligned} \quad (2)$$

From equation (2), it is pointed out that, when $n = 2$, it gives a discrete version of the analog second derivative value.



FIGURE 1: Experimental hardware for water distribution system.

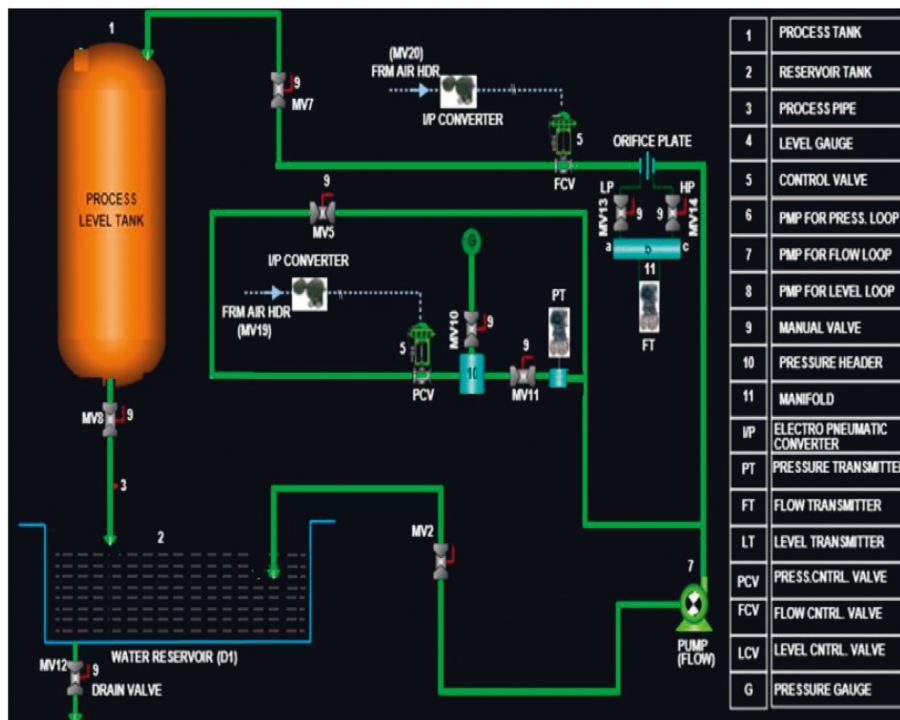


FIGURE 2: Process flow diagram for the experimental setup.

It represents the local acceleration of the time-series forecasting values instead of a normal trend.

Hence, for the future predicted value M , the forecasting formula is represented as follows:

$$M_t = \mu + \phi_1 M_{t-1} + \dots + \phi_p M_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad (3)$$

where θ is the moving average parameter, M is the predicted value, and e is the error terms sampled with zero mean. Equation (3) can be rewritten and represented as the autoregressive model equation as follows:

$$\left(1 - \sum_{i=1}^p \alpha_i L^i\right) M_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) e_t, \quad (4)$$

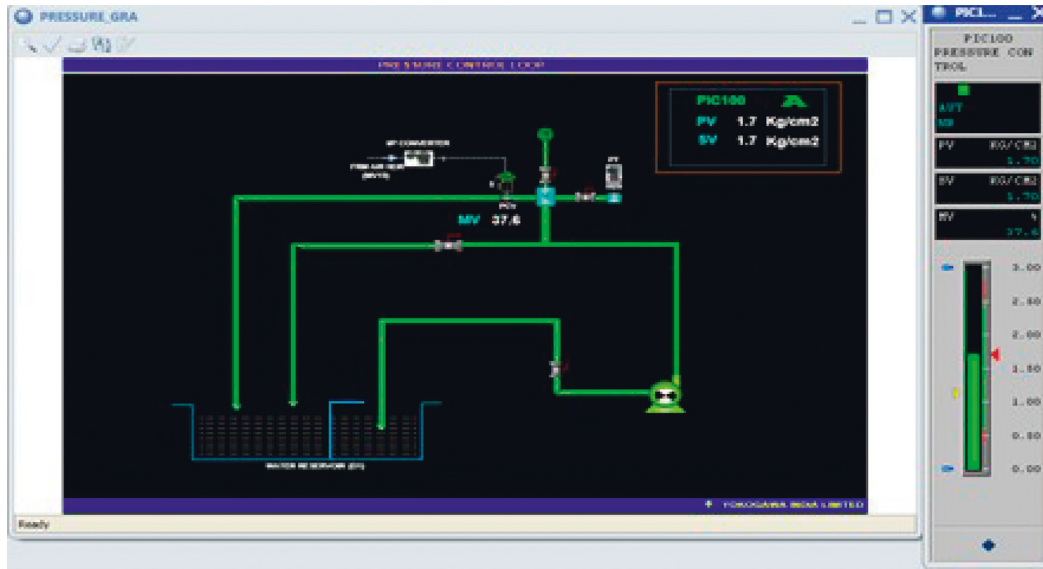


FIGURE 3: Pressure control loop for water distribution system.

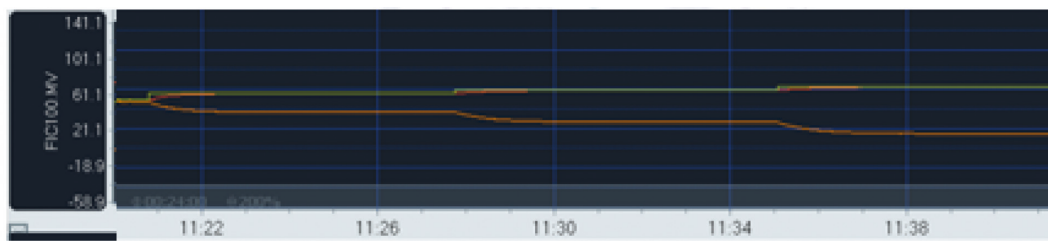


FIGURE 4: Front-end HMI for pressure data recording.

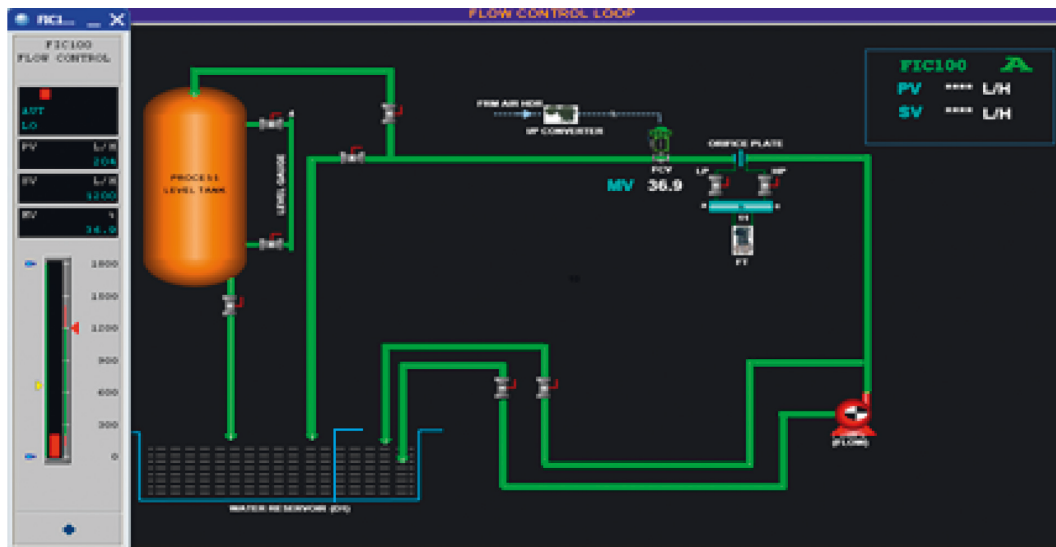


FIGURE 5: Flow control loop for water distribution system.

where L represents the lag operator, and α represents the parameters of all autoregressive part as mentioned in equation (3). An ordinary least squares called as OLS is used to for predicting the slope coefficients in the ARIMA model. The unit test root method is used for this purpose. From

equation (4), $(1 - \sum_{i=1}^p \alpha_i L^i)$ is involved with the unit root $1 - L_i$ with the multiple factor U as follows:

$$\left(1 - \sum_{i=1}^p \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p-u} \alpha_i L^i\right) (1 - L)^u. \quad (5)$$

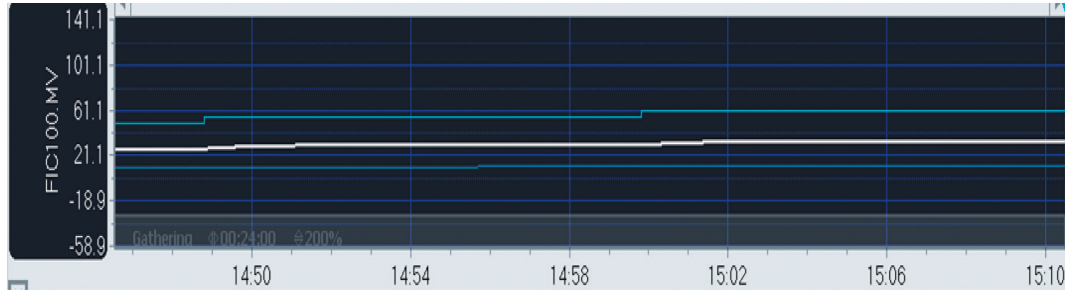


FIGURE 6: Front-end HMI for flow data recording.

The ARIMA (p, u, q) process is represented, and then, equation (4) can be rewritten based on equation (5) as follows:

$$\begin{aligned} \left(1 - \sum_{i=1}^p \alpha_i L^i\right) (1-L)^u M_t \\ = \left(1 + \sum_{i=1}^q \theta_i L^i\right) e_t. \end{aligned} \quad (6)$$

From equation (5), the generalized ARIMA equation is defined by the following term, ARIMA $(P, Q+1)$ process with “ u ” roots.

$$\begin{aligned} \left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1-L)^u M_t \\ = \delta + \left(1 - \sum_{i=1}^q \theta_i L^i\right) e_t. \end{aligned} \quad (7)$$

The following steps are involved in data processing using the ARIMA model for future prediction using R software functions.

- (i) The dataset values obtained are first imported to the environment part in R studio software and then tested for class. The class should be in a time-series format to predict and forecast future data.
- (ii) The dataset is then converted to .ts from .xlsx format and checked for autocorrelation function—`acf()`, partial autocorrelation function—`pacf()`, augmented Dickey-Fuller test `adf.test()` to check the significant stationarity of the time-series dataset values.
- (iii) The time-series dataset values are partially or completely converted into a simple or fully formed statistical model by using the ARIMA model, and the p -value is checked for the newly formed model to check the sustainability of the formed statistical model. The p -value should be less than alpha 0.05 to conclude that the model is statistically significant [9, 10].
- (iv) `acf()`, `pacf()`, and `ADF.test()` functions are used to check the newly formed ARIMA model to check the significant nature of the framed model.

- (v) The responses with the forecasted value are obtained from the R studio, and the predicted data are collected from the execution of the comments in the console part of the R studio.

3. Results

In order to predict the future needs in supplying water through a pipeline, historical output data of the water pipeline system have been regularly taken. With the help of experimental hardware, a certain amount of data has been periodically collected and those data can be used as datasets for future prediction of water requirements. Those pressure datasets and flow datasets are going to play a vital role in future predictions and forecasting [11, 12].

3.1. Forecasting the Possibility of Occurrence of Cracks and Leaks Based on Pressure Datasets

3.1.1. Input Time-Series Data Graph. The input dataset values obtained from the Excel by controlling the pressure control valve attached to the pipeline over equal intervals of time for three different conditions, namely, normal, crack, and damaged, which are fed to the R studio, are plotted and shown in Figure 7.

3.1.2. Forecasting Process-Normal Condition. The input dataset values of normal condition are imported through the SCADA system to the R studio. The plot of the values associated with the pressure at normal condition is shown in Figure 8. The input dataset values of pressure at the normal condition are fed to the R studio; as the forecasting process requires statistical data, the R studio converts the input data into a time-series data with the time-series function and the plot of the time-series dataset values associated with the pressure at the normal condition with respect to the time [42].

The stability and sustainability of the time-series data of pressure at normal condition are checked, and the results of `acf()` and `pacf()` plots are shown in Figure 9.

Figure 10 shows that the data are not within the range of expectations of the stability, so here comes the necessity of making the data into an ARIMA model to obtain statistically stable data. The ARIMA model is built for prediction; the

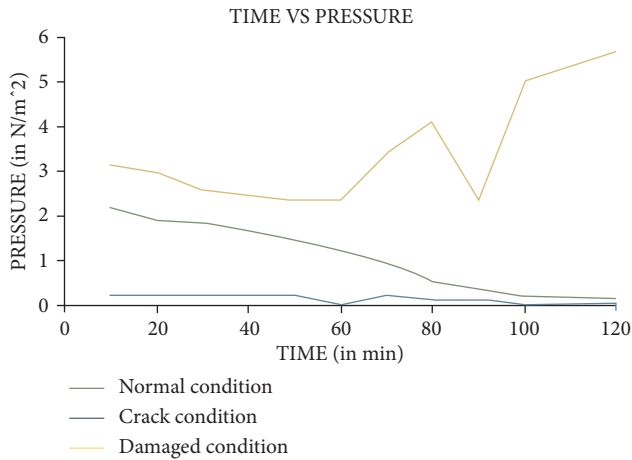


FIGURE 7: Input time-series data graph obtained from Excel.

stability and sustainability of the ARIMA model of the pressure residuals at normal condition are checked; and the results of $acf()$ and $pacf()$ plots are shown in Figure 10.

Figure 11 shows that the data are within the range of expectations of the stability, and the p -value is less than 0.05, which represents the stability of the data in the ARIMA model of pressure at normal condition over a period of time, which can be used for forecasting. The lag value represents the lead time period over which the average of moving values obtained from the preceding data values remains to be same for the same lead time period in the succeeding data. The predicted value of the ARIMA model of pressure forecast at damaged condition is plotted on the graph as shown in Figure 11.

Figure 11 represents the future forecasted data of pressure at normal condition for the next 1 hour where the blue line represents the average of the forecast value, which ranges with a minimum peak of 0.451379 N/m^2 and 2.022273 N/m^2 as the maximum peak. The shaded region represents the frequency of the forecast value with maximum peaks ranging from 1.706869 N/m^2 to 3.065836 N/m^2 and the minimum peaks ranging from -0.81046 N/m^2 to 1.042164 N/m^2 . The average of the forecast value reduces in its frequency as the pressure from the pressure control valve is reduced over equal intervals of time, and it is assumed to be constant when the pressure from the pressure control valve remains constant. The predicted time-series data are based on the pressure data collected with very less frequency say 125 minutes of cumulative data. In case if the frequency is high, say a month or early data the prediction accuracy will be more.

3.1.3. Forecasting Process-Crack Condition. The input dataset values of the crack condition are fed to the R studio, and the plot of the values associated with the pressure at the crack condition is shown in Figure 12.

The input dataset values of pressure at the normal condition in the pipe are fed to the R studio; as the forecasting process requires statistical data, the R studio converts the input data into a time-series data with the time-series

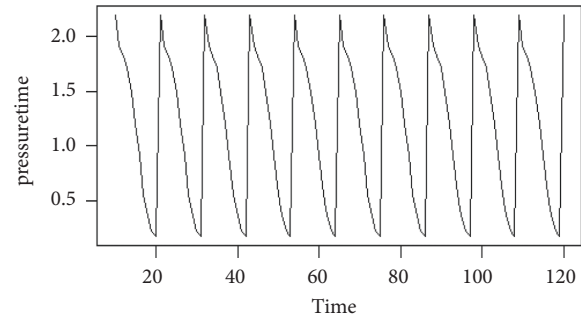


FIGURE 8: Input time-series plot.

function and the plot of the time-series dataset values associated with the pressure at the normal condition with respect to the time that is shown in Figure 13.

The stability and sustainability of the time-series data of pressure at crack condition are checked, and the results of $acf()$ and $pacf()$ plots are shown in Figure 14.

Figure 14 shows that the data are not within the range of expectations of the stability, so here comes the necessity of making the data into an ARIMA model to obtain statistically stable data. The ARIMA model is built for prediction; the stability and sustainability of the ARIMA model of the pressure residuals at crack condition are checked; and the results of $acf()$ and $pacf()$ plots are shown in Figure 15.

Figure 15 shows that the data are within the range of expectations of the stability, and the p -value is less than 0.05, which represents the stability of the data in the ARIMA model of pressure at crack condition over a period of time, which can be used for forecasting. The lag value represents the lead time period over which the average of moving values obtained from the preceding data values remains to be same for the same lead time period in the succeeding data. The predicted value of the ARIMA model of pressure forecast at crack condition is plotted on the graph as shown in Figure 16.

Figure 16 represents the future forecasted data of pressure at crack condition for the next 1 hour where the blue line represents the average of the forecast value remains same as 0.144865 N/m^2 as the pressure remains the same at the crack region without any rise or fall in pressure irrespective of the change in pressure made by the pressure control valve. The minimum peak of the forecast value remains at -0.02742 N/m^2 , and the maximum peak of the forecast value remains at 0.317153 N/m^2 . This infers that the pressure remains unchanged over a period of time till the crack is sealed in the pipeline. The estimated value seems like constant with respect to time, since less frequency pressure data are taken for the forecasting analysis. The dataset's quantitative expansion improves prediction accuracy.

3.1.4. Forecasting Process-Damaged Condition. The input dataset values of normal condition are fed to the R studio, and the plot of the values associated with the pressure at normal condition is shown in Figure 17.

The input dataset values of pressure at the normal condition in the pipe are fed to the R studio; as the

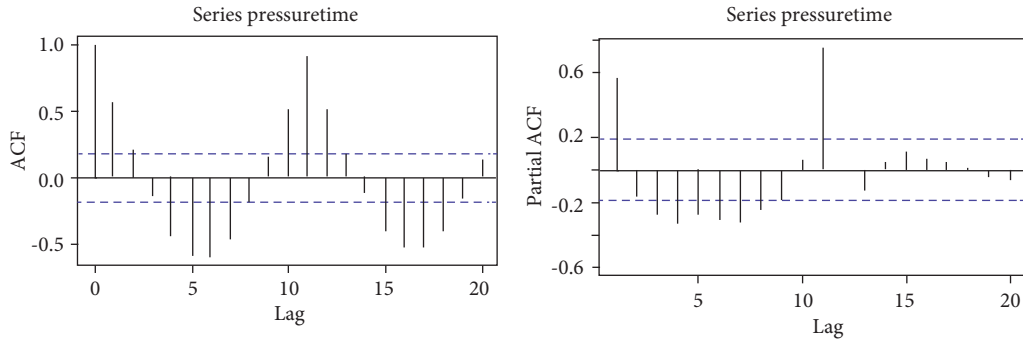


FIGURE 9: Checking stability of the data.

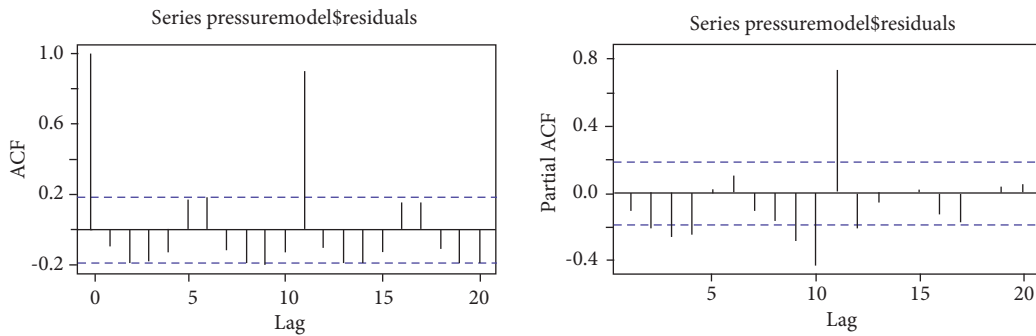


FIGURE 10: Checking of the ARIMA model data (stable).

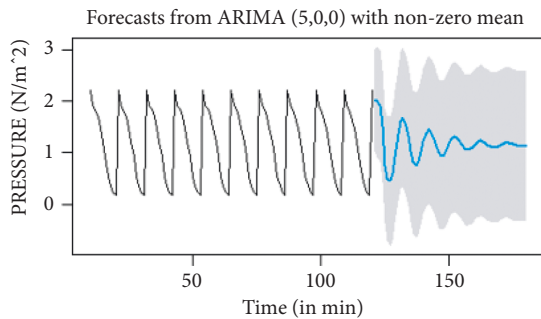


FIGURE 11: Forecast time-series data graph from R studio.

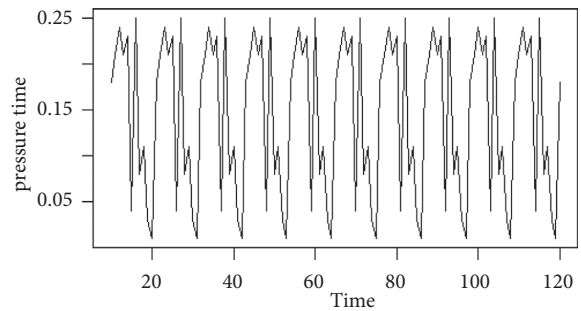


FIGURE 13: Input time-series plot.

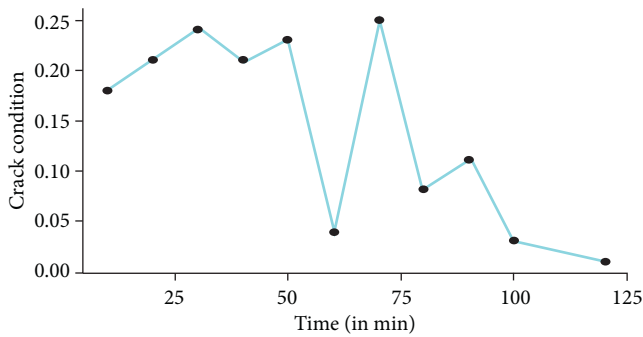


FIGURE 12: Input dataset value.

forecasting process requires a statistical data, the R studio converts the input data into a time-series data with the time-series function and the plot of the time-series dataset values

associated with the pressure at the normal condition with respect to the time that is shown in Figure 18.

The stability and sustainability of the time-series data of pressure at damaged condition are checked, and the results of `acf()` and `pacf()` plots are shown in Figure 19.

Figure 19 shows that the data are not within the range of expectations of the stability, so here comes the necessity of making the data into an ARIMA model to obtain statistically stable data. The ARIMA model is built for prediction; the stability and sustainability of the ARIMA model of the pressure residuals at damaged condition are checked, and the results of `acf()` and `pacf()` plots are shown in Figure 20.

Figure 20 shows that the data are within the range of expectations of the stability, and the p -value is less than 0.05, which represents the stability of the data in the ARIMA model of pressure at damaged condition over a period of time, which

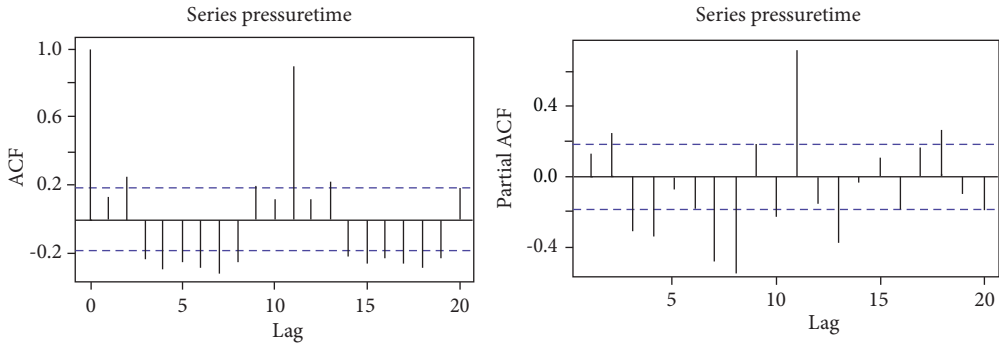


FIGURE 14: Checking stability of the data.

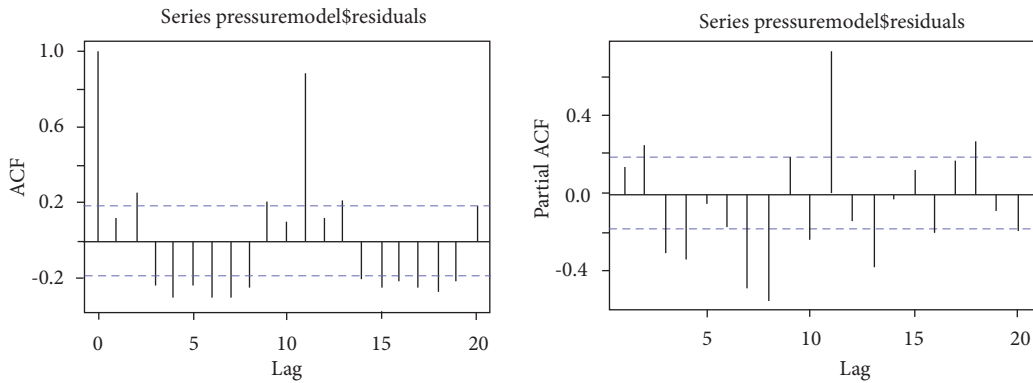


FIGURE 15: Checking of the ARIMA model data (stable).

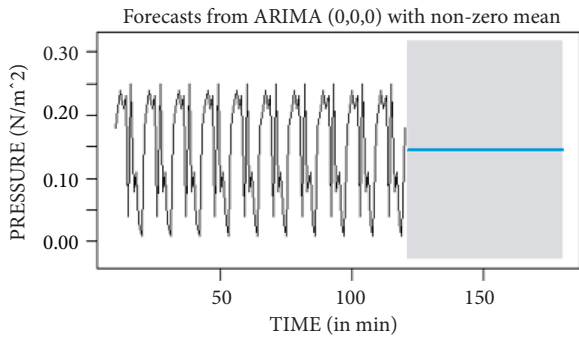


FIGURE 16: Forecast time-series data graph from R studio.

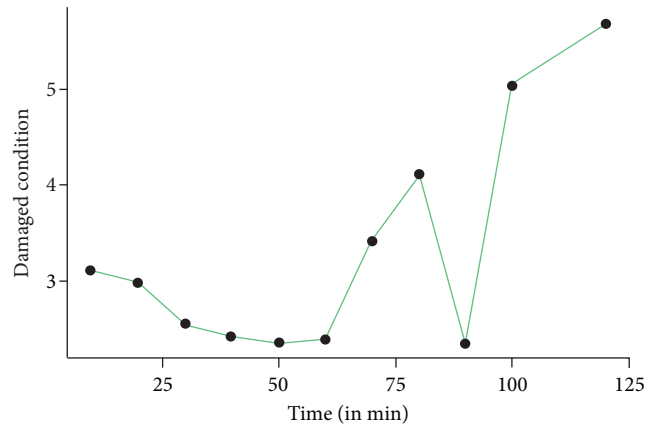


FIGURE 17: Input dataset value.

can be used for forecasting. The lag value represents the lead time period over which the average of moving values obtained from the preceding data values remains to be same for the same lead time period in the succeeding data. The predicted value of the ARIMA model of pressure forecast at damaged condition is plotted on the graph as shown in Figure 21.

Figure 21 represents the future forecasted data of pressure at damaged condition for the next 1 hour where the blue line represents the average of the forecast value, which increases from the minimum peak of 2.880788 N/m^2 to the maximum peak of 3.29797 N/m^2 and becomes constant. The shaded region represents the frequency of the forecast value with maximum peaks ranging from 4.866227 N/m^2 to 5.664348 N/m^2 .

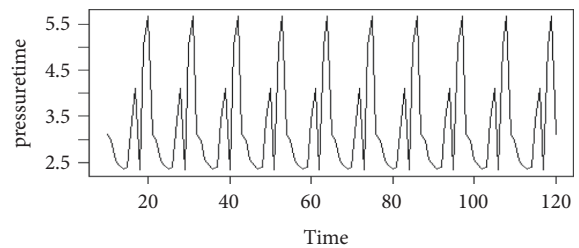


FIGURE 18: Input time-series plot.

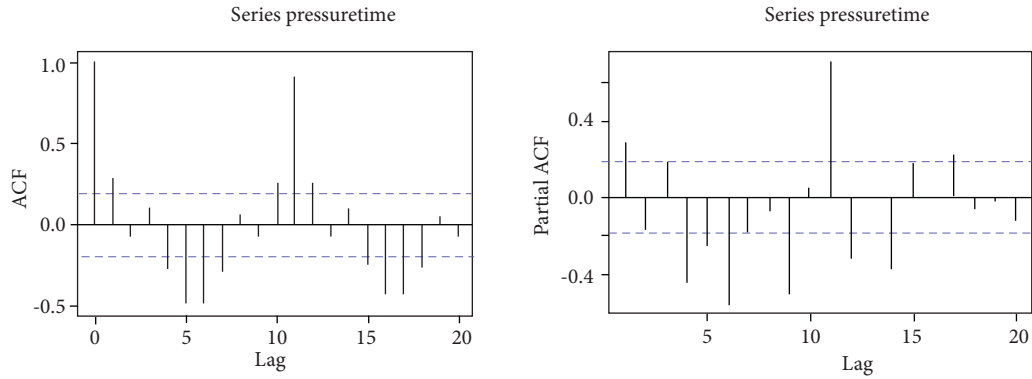


FIGURE 19: Checking stability of the data.

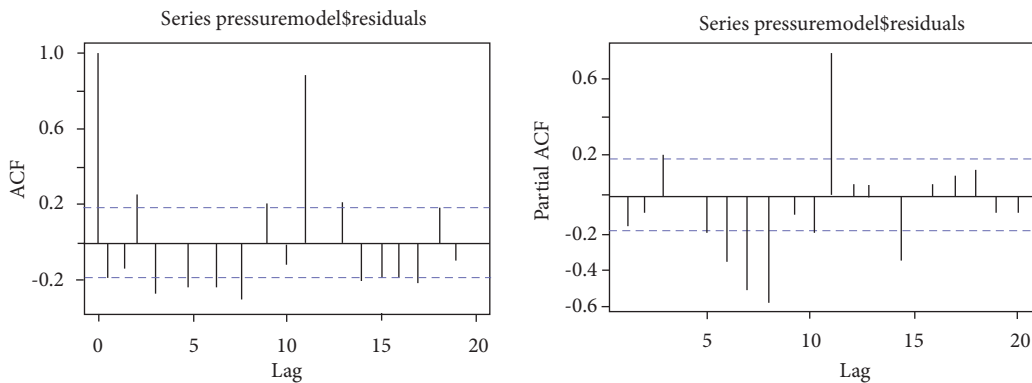


FIGURE 20: Checking of the ARIMA model data (stable).

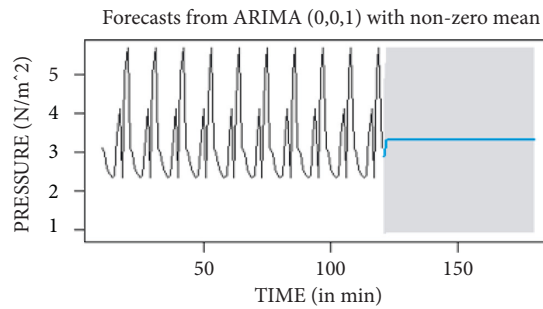


FIGURE 21: Forecast time-series data graph from R studio.

m^2 and the minimum peaks ranging from $0.895349 N/m^2$ to $0.931593 N/m^2$. The constant average value of the forecast infers that the pressure remains unchanged over a period of time till the damaged pipeline is replaced with a new pipeline.

3.2. Flow Forecasting

3.2.1. *Input Time-Series Data Graph.* The input dataset values obtained from Excel by monitoring the flow over a pipeline on a weekly basis, which are fed to the R studio, are plotted and shown in Figure 22.

3.2.2. *Forecasting Process.* The input dataset values of flow in the pipe are fed to the R studio; as the forecasting process

requires a statistical data; the R studio converts the input data into a time-series data with the time-series function; and the plot of the time-series dataset values associated with the flow with respect to the time that is shown in Figure 23.

The stability and sustainability of the time-series data of flow are checked, and the results of `acf()` and `pacf()` plots are shown in Figure 24. The data are not within the range of expectations of the stability, so here comes the necessity of making the data into an ARIMA model to obtain statistically stable data.

Figure 24 shows that the data are not within the range of expectations of the stability, so here comes the necessity of making the data into an ARIMA model to obtain statistically stable data. The ARIMA model is built for prediction; the

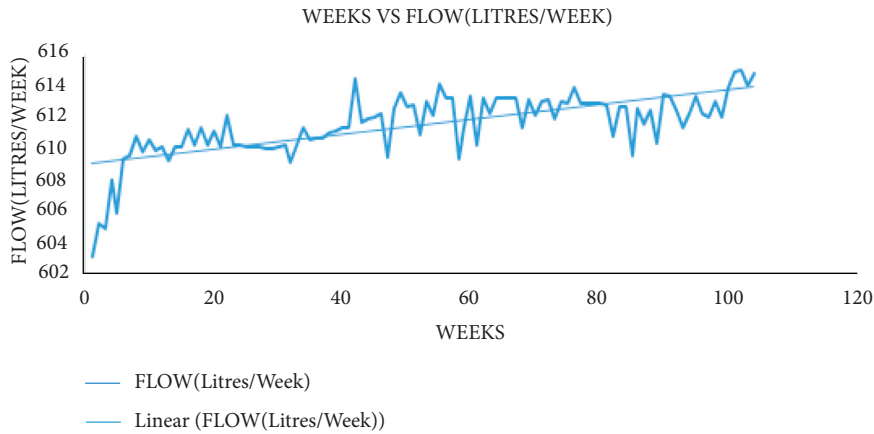


FIGURE 22: Input dataset value graph.

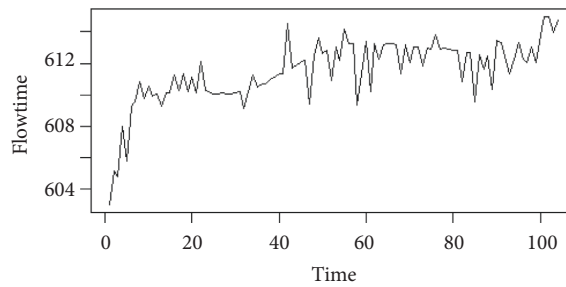


FIGURE 23: Input time-series plot.

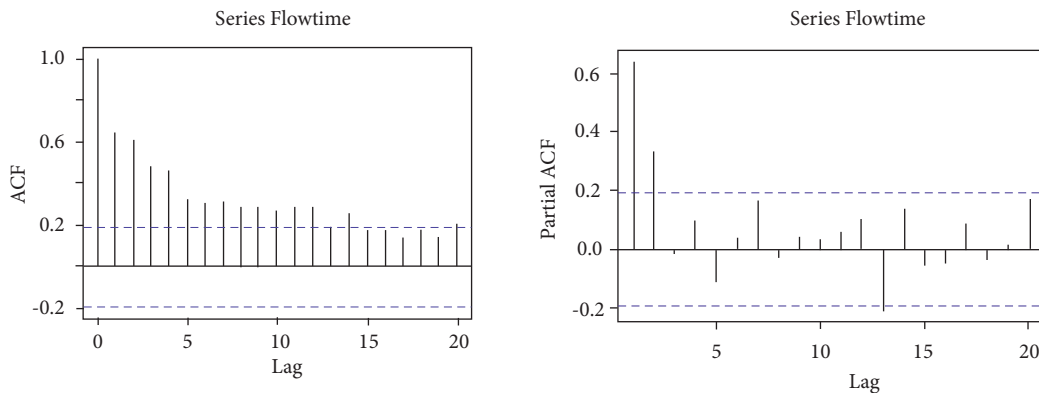


FIGURE 24: Checking stability of the data.

stability and sustainability of the ARIMA model of the flow residuals are checked, and the results of `acf()` and `pacf()` plots are shown in Figure 25.

Figure 25 shows that the data are within the range of expectations of the stability, and the p -value is less than 0.05, which represents the stability of the data in the ARIMA model of flow over a period of time, which can be used for forecasting. The lag value represents the lead time period over which the average of moving values obtained from the preceding data values remains to be same for the same lead time period in the succeeding data. The predicted value of the ARIMA model of flow forecast is plotted on the graph as shown in Figure 26.

Figure 26 represents the future forecasted data of water requirement for the next 1 year where the blue line represents the average of the forecast value, which increases from the minimum peak of 614.6292 (liters/week) to the maximum peak of 620.0099 (liters/week) and becomes constant. The shaded region represents the frequency of the forecast value with maximum peaks ranging from 617.0086 (liters/week) to 628.5465 (liters/week) and the minimum peaks ranging from 611.0967 (liters/week) to 612.2914 (liters/week). This infers that the requirement of water quantity in the next year increases as represented by the average of the forecast data.

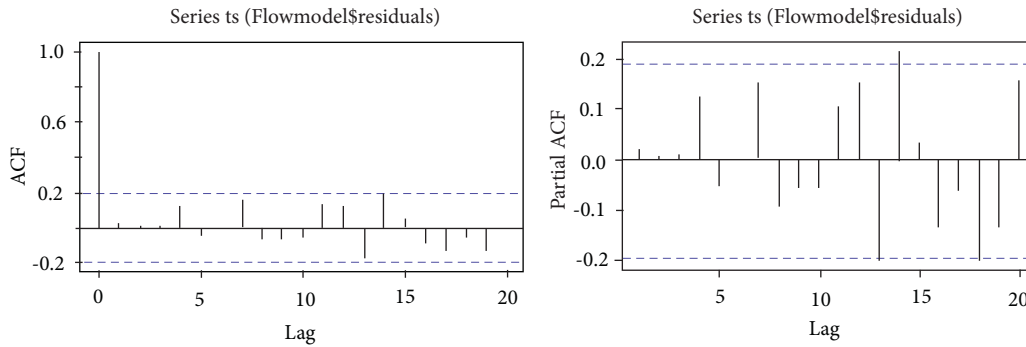


FIGURE 25: Checking of the ARIMA model data (stable).

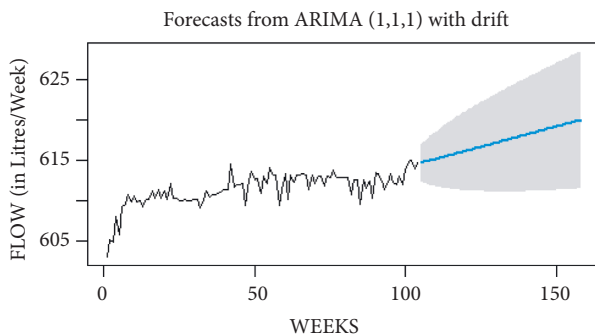


FIGURE 26: Forecast time-series data graph from R studio.

4. Conclusions

The identification of the crack, pressure difference, leakages, and blockages with the help of abnormal pressure rise or drop in the municipal water distribution system was analyzed [43, 44]. On the other side, the water flow in the pipeline was monitored and prediction for the future requirements was made by analyzing its present utilization of the flow datasets obtained from the preceding two years. With the help of the time-series forecasting algorithm, the collection of datasets for the flow periodically was made [45, 46]. Those datasets have been used as a source of input to the R studio under the time-series forecasting algorithm. The flow rate datasets of the past two years were recorded, and the datasets are analyzed on the time-series forecasting algorithm. As an output of the algorithm, the need for the next year was predicted and figured out in the form of a graph. This provides a strategy for forecasting the needs for the upcoming year. The stability of the statistical data was confirmed by using the p -value (less than 0.05) provided by the ARIMA model of data for both pressures and flow datasets. The future prediction of the pressure concerning the flow in damaged conditions would range between 2.880788 and 3.29797 N/m². The future prediction of the pressure concerning flow in normal conditions would range between 0.451379 and 2.022273 N/m². The future prediction of the pressure concerning flow in crack condition would remain unchanged but lower than the normal condition. The future prediction of the flow requirement would range between 614.6292 and 620.0099 liters/week. Furthermore, future studies may be incorporated based

on the smart pig for the identification of fault and data transmission in oil pipelines.

Data Availability

The required data can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Editing and writing were carried out at the Debre Tabor University, Ethiopia.

References

- [1] A. Terkawi and N. Innab, "Major impacts of key reinstallation attack on internet of things system," in *Proceedings of the 2018 21st Saudi Computer Society National Computer Conference (NCC)*, pp. 1–6, Riyadh, Saudi Arabia, April 2018.
- [2] H. M. Ramos, A. McNabola, P. A. López-Jiménez, and M. Pérez-Sánchez, "Smart water management towards future water sustainable networks," *Water*, vol. 12, no. 1, p. 58, 2019.
- [3] X. Wei, L. Zhang, H.-Q. Yang, L. Zhang, and Y.-P. Yao, "Machine Learning for Pore-Water Pressure Time-Series Prediction: Application of Recurrent Neural Networks," *Geoscience Frontiers*, vol. 12, no. 1, pp. 453–467, 2021.
- [4] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: a review," *Sensors*, vol. 18, no. 8, Article ID 2674, 2018.
- [5] N. Y. K. Phan and H. T. Nguyen, "Inflammatory bowel disease classification improvement with metagenomic data binning using mean-shift clustering," in *Communications in Computer and Information Science (CCIS)*, T. K. Dang, J. Küng, M. Takizawa, and T. M. Chung, Eds., FDSE 2020, Springer, Singapore, pp. 294–308, 2020.
- [6] G. Liu, Ya Zhang, W.-J. Knibbe et al., "Potential impacts of changing supply-water quality on drinking water distribution," *International Journal on ELSEVIER*, vol. 116, 2020.
- [7] M. M. Giraldo-González and J. P. Rodríguez, "Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks," *Water*, vol. 12, no. 4, 2020, <https://www.mdpi.com/about/journals>.
- [8] Q. Fang, J. Zhang, C. Xie, and Y. Yang, "Detection of multiple leakage points in water distribution networks based on

- convolutional neural networks,” *Water Supply*, vol. 19, no. 8, pp. 2231–2239, 2019.
- [9] N. A. Kumar, G. Shyni, G. Peter, A. A. Stonier, and V. Ganji, “Architecture of network-on-chip (NoC) for secure data routing using 4-H function of improved TACIT security algorithm,” *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–9, Article ID 4737569, 2022.
- [10] E. B. Priyanka, S. Thangavel, V. Madhuvishal, S. Tharun, K. V. Raagul, and C. S. S. Krishnan, “Application of integrated IoT framework to water pipeline transportation system in smart cities,” *Intelligence in Big Data Technologies—Beyond the Hype*, Springer, pp. 571–579, Singapore.
- [11] G. Peter, J. Livin, and A. Sherine, “Hybrid optimization algorithm based optimal resource allocation for cooperative cognitive radio network,” *Array*, vol. 12, Article ID 100093, 2021.
- [12] E. B. Priyanka, C. Maheswari, S. Thangavel, and M. P. BalaBala, “Integrating IoT with LQR-PID controller for online surveillance and control of flow and pressure in fluid transportation system,” *Journal of Industrial Information Integration*, vol. 17, Article ID 100127, 2020.
- [13] J. S. Armstrong, *Long-range Forecasting: From crystal ball to Computer*, John Wiley & Sons, NJ, USA, 1978.
- [14] S. Wambura, J. Huang, and H. Li, “Long-range forecasting in feature-evolving data streams,” *Knowledge-Based Systems*, vol. 206, 2020.
- [15] S. Mehran Jafari, A. Reza Zahiri, O. Bozorg Hadad, and M. M. R. Tabari, “A hybrid of six soft models based on ANFIS for pipe failure rate forecasting and uncertainty analysis: a case study of gorgan city water distribution network, methodologies and application,” *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 25, no. 11, pp. 7459–7478, 2021.
- [16] J. Wang, F. Zhou, L. Zhang et al., “Experimental study of wax deposition pattern concerning deep condensate gas in Bozi block of Tarim Oilfield and its application,” *Thermochimica Acta*, vol. 671, pp. 1–9, 2019.
- [17] M. Monteiro and M. Costa, “A time series model comparison for monitoring and forecasting water quality variables,” *Hydrology*, vol. 5, no. 3, p. 37, 2018.
- [18] J. F. Joe and T. V. Raja, “Enhanced robustness for digital images using geometric attack simulation,” *Procedia Engineering*, vol. 38, pp. 2672–2678, 2012, <https://doi.org/10.1016/j.proeng.2012.06.314>.
- [19] J. Liu, S. Wang, N. Wei, Xi Chen, H. Xie, and J. Wang, “Natural gas consumption forecasting: a discussion on forecasting history and future challenges,” *Journal of Natural Gas Science and Engineering*, vol. 90, Article ID 103930, 2021.
- [20] A. O. Adejo, A. J. Onumanyi, J. M. Anyanya, and S. O. Oyewobi, “Oil and gas process monitoring through wireless sensor networks: a survey,” *Journal of Applied Sciences*, vol. 6, pp. 5660–5568, 2013.
- [21] G. Dundulis, A. Grybenas, R. Karalevicius, and V. Makarevicius, “Static analytical and experimental research of energy analysis to safeguard the oil transportation system,” *Energy analysis statistics*, vol. 67, pp. 239–248, 2018.
- [22] P. Hopkins, “The structural integrity of oil and gas transmission pipelines,” *Computer Structural Integrity*, vol. 1, pp. 87–123, 2018.
- [23] Y. Chen, L. Song, Y. Liu, L. Yang, and D. Li, “A review of the artificial neural network models for water quality prediction,” *Applied Sciences*, vol. 10, no. 17, p. 5776, 2020.
- [24] D. P. Abreu, K. VelasquezVelasquez, M. CuradoCurado, and E. MonteiroMonteiro, “A resilient Internet of Things architecture for smart cities,” *Annals of Telecommunications*, vol. 72, no. 1–2, pp. 19–30, 2017.
- [25] M. A. Khan and K. Salah, “IoT security: review, blockchain solutions, and open challenges,” *Future Generation Computer Systems*, vol. 82, pp. 395–411, 2018.
- [26] M. Kim, “A quality model for evaluating IoT applications,” *International Journal of Computer and Electrical Engineering*, vol. 8, no. 1, pp. 66–76, 2016.
- [27] O. Novo, “Blockchain meets IoT: an architecture for scalable access management in IoT,” *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1184–1195, 2018.
- [28] S. Susan Jacob and Dr. R. Vijayakumar, “Sentiment analysis over twitter bigdata using modified meanshift clustering,” vol. 7, no. 15, pp. 5835–5841, 2020.
- [29] G. Bin Iderus and S. Bin Iderus, “Design of enhanced energy meter using GSM prepaid system and protective relays,” *Materials Today Proceedings*, vol. 39, pp. 582–589, 2021.
- [30] R. Ranjbarzadeh and B. Saadi, “Automated liver and tumor segmentation based on concave and convex points using fuzzy c-means and mean shift clustering,” *Measurement*, vol. 150, Article ID 107086, 2020.
- [31] D. V. Brower and N. Prescott, “Real-time subsea monitoring and control smart field solutions,” *Subsea Rio 2018 Conference*, vol. 3, pp. 3413–3462, 2018.
- [32] J. Najafi, A. Peiravi, A. Anvari-Moghaddam, and J. M. Guerrero, “An efficient interactive framework for improving resilience of power-water distribution systems with multiple privately-owned microgrids,” *International Journal of Electrical Power & Energy Systems*, vol. 116, Article ID 105550, 2020.
- [33] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, NJ, USA, 1994.
- [34] M. Khashei and M. Bijari, “A novel hybridization of artificial neural networks and ARIMA models for time series forecasting,” *Applied Soft Computing*, vol. 11, no. 2, pp. 2664–2675, 2011.
- [35] V. A. O. P. da Silva, R. C. de Freitas, P. R. de Oliveira et al., “Microfluidic paper-based device integrated with smartphone for point-of-use colorimetric monitoring of water quality index,” *Measurement*, vol. 164, Article ID 108085, 2020.
- [36] T. M. P. Junaidi, A. Surtono, G. A. Puazi, S. W. Suciyan, I. Firdaus, and L. Rumiyan, “Measurement of physical parameters of water quality in real-time based on arduino,” *Journal of Physics: Conference Series*, vol. 1751, no. 1, Article ID 012067, 2021.
- [37] R. I. Pereira, I. M. Dupont, P. C. Carvalho, and S. C. Jucá, “IoT embedded Linux system based on Raspberry Pi applied to real-time cloud monitoring of a decentralized photovoltaic plant,” *Measurement*, vol. 114, pp. 286–297, 2018.
- [38] E. B. Priyanka, C. Maheswari, and B. Meenakshipriya, “Parameter monitoring and control during petrol transportation using PLC based PID controller,” *Journal of Applied Research and Technology*, vol. 14, no. 2, pp. 125–131, 2016.
- [39] G. Sherine and A. Sherine, “Induced over voltage test on transformers using enhanced Z-source inverter based circuit,” *Journal of Electrical Engineering*, vol. 68, no. 5, pp. 378–383, Sep. 2017.
- [40] S. Bhoumik, S. Chatterjee, A. Sarkar, A. Kumar, and F. J. John Joseph, “Covid 19 prediction from X ray images using fully connected convolutional neural network,” in *Proceedings of the CSBio '20: Proceedings of the Eleventh International Conference on Computational Systems-Biology and Bioinformatics*, pp. 106–107, ACM, Bangkok Thailand, November 2020.

- [41] G. Petris, C. Patrizia, and P. Sonia, *Dynamic Linear Models with R*, Springer Science & Business Media, Germany, 2009.
- [42] K. L. Keung, C. K. M. Lee, K. K. H. Yeung, and C. K. Yeung, "Smart city application and analysis: real-time urban drainage monitoring by IoT sensors: a case study of Hong Kong," in *Proceedings of the 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 521–525, Bangkok, Thailand, December 2018.
- [43] Z. Jia, Z. Wang, W. Sun, and Z. Li, "Pipeline leakage localization based on distributed FBG hoop strain measurements and support vector machine," *Optik*, vol. 176, pp. 1–13, 2019.
- [44] J. Shen, Z. Gui, S. Ji, J. Shen, H. Tan, and Y. Tang, "Cloud-aided lightweight certificateless authentication protocol with anonymity for wireless body area networks," *Journal of Network and Computer Applications*, vol. 106, pp. 117–123, 2018.
- [45] A. Y. Scanlon, B. R. Scanlon, "How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions," *Environmental Research Letters*, vol. 14, no. 7, p. 073001, 2019.
- [46] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," *Future Generation Computer Systems*, vol. 82, pp. 375–387, 2018.