Detection and Prediction of Leakages in Water Distribution Networks

Mariaelena Berlotti[1][a], Sarah Di Grande[1][b], Salvatore Cavalieri[1][c] and Roberto Gueli[2][d]
[1]*Department of Electrical Electronic and Computer Engineering, University of Catania, Viale A. Doria n.6, Catania, Italy*
[2]*EHT, Viale Africa n.31, Catania, Italy*

Keywords: Water Distribution Systems, Leakages, Anomaly Detection, Predictive Maintenance.

Abstract: Leakages are one of the main causes of water loss in a water distribution system (WDS). In recent years, the increasing of streaming data coming from sensors installed in the water network, allows the monitoring the health status of each asset of the WDS. In this paper, a preliminary data-driven approach for leakages detection and prediction is proposed. Starting from the characteristics of a real water distribution network, a realistic leakages dataset has been achieved. Using this dataset, unsupervised rule-based time series algorithms has been trained for the detection and prediction of leakages.

1 INTRODUCTION

Water loss in the Water Distribution Systems (WDSs) is a topic that has been attracted much attention in recent years. The International Water Association (IWA) defines leakages as an important form of loss of water from a WDS due to leaks (Pearson, 2019). Leak is considered as a failure causing an unplanned loss of water from a network. The term is generic and can be used to define leaks of any size and referred to any type of asset from pipes and valves to reservoirs.

As concerns the sizes, leaks can be categorized as *abrupt* leakages and *incipient* leakages (Vrachimis et al., 2018). Abrupt leaks are leakages that occurs suddenly in a water system and results in large volumes of water coming out of the network in a short period; generally, this type of leakages is associated to pipe burst. Incipient leakages, instead, increase gradually over time starting as background leakages and developing into full-blown leakages (Tornyeviadzi and Seidu, 2023).

Detection of incipient leakages is difficult as they typically occur in pipes with smaller diameter, casusing at the beginning low volumes of water loss; this type of leaks growth slowly during the time leading to huge losses if not discovered and repaired in due time (Wan et al., 2022).

Leakage detection is defined as the process of locating and pinpointing water leaks (NAIADES, 2022).

Among the existing approaches for leakages detection there are those based on data-driven models. These models rely on learning techniques applied on a collection of data coming from the WDS and for this reason, they do not require a domain knowledge about the network. On the other hand, a large amount of historical data is needed to perform the analysis (Escofet et al., 2016).

Prediction of water leakages involves spotting leaks before they happen (Cody, 2020). Leakages prediction methods are used to identify areas and pipes in the network with a high probability of leakage, allowing water utilities to create an appropriate active leakage control plan (Leu and Bui, 2016). Prediction of water leak is a challenging task: tightness and invisibility of the hydraulic components as the rarity and uncertainty of these events makes the prediction of these faults events difficult (Wang et al., 2022).

Data-driven approaches have emerged as a powerful tool for predictive maintenance applications; indeed, the increase of data availability collected through the sensors and the smart meters lead to the beginning of a digitalization process of the

[a] https://orcid.org/0009-0008-8895-2175
[b] https://orcid.org/0009-0007-6564-704X
[c] https://orcid.org/0000-0001-9077-3688
[d] https://orcid.org/0000-0002-8014-0243

Berlotti, M., Di Grande, S., Cavalieri, S. and Gueli, R.
Detection and Prediction of Leakages in Water Distribution Networks.
DOI: 10.5220/0012122000003541
In *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, pages 436-443
ISBN: 978-989-758-664-4; ISSN: 2184-285X

water sector, known as Water 4.0 (Adedeji et al. 2022). Water 4.0 includes the service innovation of water networks: maintenance becomes preventive and predictive programmed on the basis of signals data (Caldognetto et al., 2022).

In this paper, the authors are going to propose a data-driven approach to realize both detection and prediction of leakages, considering a real WDS of the city of Milan (Italy). The machine learning models aimed to realize this aim have been defined. Using these models, a performance evaluation has been carried out, comparing the leaks detected with the real ones. Finally, the selected algorithm has been used for leakages prediction and the relevant results have been achieved.

The paper is structured as follows: Section 2 will give an overview about the similar approaches found in the literature, in order to point out the originality of the work proposed. In Section 3 the authors will give a description of the approach proposed for detecting and predicting leakages. Section 4 presents the results obtained from the application of the approach on the leakages dataset. A final section will summarize the conclusions.

2 RELATED WORK

The aim of this section is to give an overview of the main approaches present in the current literature about the use of data-driven approaches for leakages detection and prediction in WDSs. This overview will allow to point out the originality of the proposal.

Leakage detection methods can be broadly classified into hardware methods and software-based methods.

Hardware methods can be further categorized into passive and active systems; if the former requires vision and sensor utilization, the latter involves the analysis of acoustic, vibration, flow or pressure signals (Chan et al, 2018). In (Hunaidi et al. 1999), leak identification through the acoustical signals given by plastic pipes is presented. More recently, (Cody et al. 2020) proposed a mixed approach where deep learning is involved for the monitoring of hydroacoustic spectrograms to pinpoint leaks on pipelines. Finally, in (Wang et al. 2021) the authors investigated the characteristics of acoustic signals obtained by simulating leaks through an experimental platform; then these signals have been passed to an artificial neural network model for leak detection.

As concerns vibration signals, (Bentoumi et al. 2017) proposed a leak-detection model based on the 'Haar' continuous wavelet; the algorithm takes as input vibration signals issued from a water pipeline and decides if there is or not a leak in the network. In (Yu et al. 2023), the authors presented a machine learning models for leak detection on vibration signals collected by wireless piezoelectric accelerometers placed in real complex water distribution systems.

Active systems comprise transient-based approaches, hydraulic model-based approaches and data-driven approaches. The core idea of transient-based approaches is that any change in the physical structure of the pipe can alter flow and pressure measurement of a system (Wan et al., 2022). To capture transient behavior, this type of analysis requires big amount of data with high sampling frequency that results in costly and too complex process. For this reason, transient approaches are not recommended for real-time monitoring of large WDSs (Colombo, 2009).

Hydraulic model-based approach instead uses mathematical functions and formulas to replicate the operation of a network. Apart from requiring domain knowledge to be built, these models need the availability of large amount of historical data for calibration. Another major drawback is that model-based methods assume WDS conditions stability over time; this is not true in real life scenarios since factors as pipe ages and roughness coefficient as time goes on increase and become increasingly influential on leak occurrence (Perez et al., 2014).

Active systems include data-driven models; in particular, three types of approaches can be actually used for the detection of leakages: Supervised, Semi-supervised and Unsupervised learning.

In Supervised learning methods binary or multi-class classifier are trained using normal and abnormal labeled data. This type of methods is rarely used for leakages detection in reality due to the lack of labeled hydraulic data. Moreover, if they let us reach high accuracy for the identification of the leaks in small and simple WDS, this is not the case for larger and more complex network (Kammoun et al., 2022).

Semi-supervised learning requires only the availability of normal labelled data; they have been adopted for water quality applications (Barros, 2023).

Finally, unsupervised learning algorithms do not rely on either normal or abnormal labeled data availability. They are widely used in the field of leakages detection since they are more flexible and realistically feasible (Kammoun et al., 2022). In particular, in this paper, the authors applied an unsupervised RNN model for leaks detection and localization on flow and pressure data coming from different realist water demands scenarios of the Leak

DB dataset. In the NAIADES project an unsupervised temporal and spatial anomaly detection approach is applied to detect leakages on pressure and water flow data of the Braila districts (NAIADES, 2022). Although many studies have investigated over the problem of the detection of the leakages in the water networks, a very limited amount of researches focused on prediction of leakages (Leu et al, 2016). In the paper (Lijuan et al., 2012), the authors presented a pipe leakages prediction approach based on a radial basis function (RBF) neural network; specifically, the authors analyze all the possible factors influencing leaks and the possible relationship existing between them that could facilitate in predicting leakages more effectively. In their work, (Leu and Bui, 2016) used a Bayesian network learning (BNL) model with an updated failure probability of each asset for leakages prediction. Finally, in (Wang et al., 2022) the authors proposed a five-dimensions digital twin model for both fault diagnosis and predictive maintenance on hydraulic system; to illustrate the effectiveness of their method, they applied it to an hydraulic cylinder.

In this paper, the authors use the unsupervised-based time series anomaly detection algorithms for leakages detection and prediction. Differently from the existing literature works, only one variable is used to perform these tasks. Indeed, the core idea proposed in the paper is to let the algorithms learn leakages changes in the pressure nodes during the anomaly detection step and to use then this information also for the prediction. The advantage of this proposal is that a reduced set of information is needed for the detection and prediction of leakages, simplifying the approach. The paper aims to provide a contribution to the current literature concerning water leakages prediction of water, considering the scarcity of research available on this topic. Furthermore, the proposed method for identifying and predicting leakages is applicable to both incipient and abrupt leakages. The authors believe that this additional aspect should be considered when evaluating the paper's contribution to the knowledge of predicting and detecting water leaks, as current literature primarily focuses on the detection of abrupt leaks.

3 APPROACH

In this paper we propose a two phases approach: leakages detection and the leakages prediction. First phase consists of the application of unsupervised algorithms for the identification of leaks events. Then, among the algorithms applied for detection we choose the best performing one for prediction.

The analysis is divided in several steps: Data Acquisition (Step 1), Data Pre-processing and Transformations (Step 2), Leak Detection (Step 3) and Leakages Prediction (Step 4). The Leakages Prediction includes a Feature Engineering step, as explained later.

In the next sections, each phase of the analysis will be described.

3.1 Data Generation

Data plays a strategic role in a machine learning approach, as known. For the problem of leakages detection and prediction in WDSs data about real losses is needed. Availability of data relevant to losses is very difficult as many time data is missing; this happens for different reasons, among which there is the lack of digital support systems to store the maintenance activities inside the water distribution system assets. To solve this problem, in the present paper the data needed to run the machine learning-based solution was synthetically generated.

In details, data was created using the Water Network Tool for Resilience (WNTR), a Python package designed to simulate and analyse resilience of water distribution networks (Klise, 2018). Simulations data is related to the actual water distribution network of Milan, Italy. For the analysis, we consider a reduced version of the original urban water network, obtained through a skeletonization process that allows us to remove those pipes and nodes that have a minimum impact on the system behaviour.

Figure 1 and Figure 2 show a planimetry of the Milan network and of its reduced version.

The analysed WDS is made by: 12,354 nodes, 17,548 pipes, 26 pumping stations, 95 booster pumps. To simulate the behaviour of the system in different days of the week, the coefficients of variation of water demand should be taken into account. To get these coefficients, the authors adopted the following approach. First of all, the real water demand coefficients recorded by the Supervised Control and Data Acquisition system (SCADA) each minute on a particular day, have been considered. SCADA are industrial applications for the control and monitoring of assets (either machines or single components of an equipment).

Then, the original coefficients have been aggregated to half an hour using the mean as aggregating function. Next, the results of this 30-minutes aggregation have been stored, summed to a random value between + 1.5 and – 1.5 and multiplied for a percentage (0.05). This percentage corresponds to a noise factor that lets us reproduce possible fluctuations of the water demand curve without altering the daily patterns of water consumption.

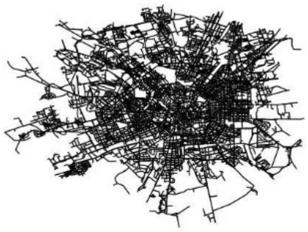

Figure 1: Milan WDS.

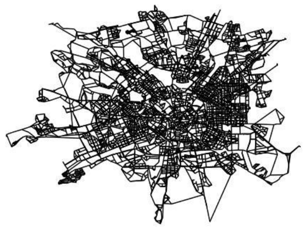

Figure 2: Skeletonized Milan WDS.

This method is repeated to generate the coefficient of variation of the half an hour water demand for the other days. These coefficients are given as input to the simulation tool to obtain hydraulics data.

Figure 3 shows the obtained water curves demand.

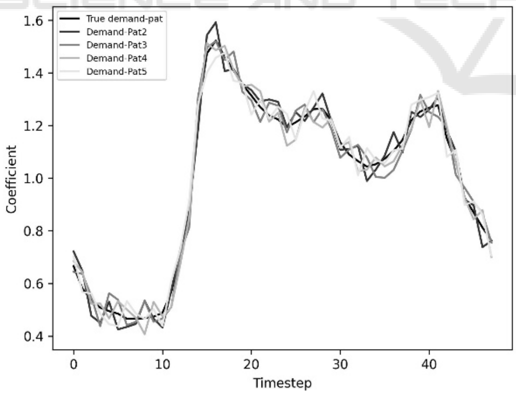

Figure 3: Water demand curves.

In Figure 3, the black-line curve represents the original water demand curve, obtained starting from the real-world coefficients. The other curves in the plot (Demand-Pat2, Demand Pat3, Demand Pat4 and Demand Pat5) are the simulated water demand curves.

WNTR package includes the possibility to add leaks in the water system.

Leakages are simulated at network nodes randomly selected. The leakage magnitude varies due to the assignment of a random leakage hole diameter. Considering the data obtained through simulations, the authors used the pressure data recorded each 30 minutes for the nodes of the skeletonized network, and the leak history used to evaluate the performances of the algorithms.

3.2 Data Pre-Processing

Data pre-processing was needed before proceeding towards the detection and prediction of the leakages.

To realize simulated leakages the WNTR simulator adopts the following method: first it divides the randomly selected pipe in two parts. Then, it adds a new node where we want to locate the leak. For this reason, duplicate columns will be present on the data achieved by the simulation: one containing the normal conditions pressure values, and the other showing a decrease in pressure data when the leak occurs (e.g., "N04755" column and "N04755_leak_node" column). We retain columns whose values reflect the occurrence of a leak. For the pressure and the leak history datasets, the time given in seconds was converted in a date format.

Table 1 and Table 2 show the final datasets after transformations.

Table 1: Pressure dataset.

Timestamp	Abbiategrasso	Anfossi	...
2009-11-18 00:00:00	66.4090	64.3000	...
2009-11-18 00:30:00	66.6450	64.3740	...
2009-11-18 01:00:00	62.2259	63.9548	...
...	...	...	...
2009-11-22 23:00:00	67.6521	64.6836	...

Table 2: Leak history dataset.

End Node	Start Time	Diameter
N03185	2009-11-18 01:30:00	0.3717
N22998	2009-11-18 16:00:00	0.1328
...	...	...
N01352	2009-11-21 02:30:00	1.2112
N10174	2009-11-22 04:00:00	0.1174

Table 1 contains a total of 12,355 columns: first column represents the Timestamp while the remaining ones, named as the nodes of the Milan network, contain 30 minutes-pressure values for each of the nodes. Table 2 is made by three columns: the 'End Node' column containing the names of the leak-nodes, the 'Start Time' column containing the time at which the leak happens, and the 'Diameter' column giving us information about the size of the leak in the pipe (in meters). Pressure data was standardized using the Standard Scaler function in order to normalize features by removing the mean and scaling to unit variance (Scikit-learn). The proposed choice is based on the fact that the hydraulic features look like standard normally distributed data.

3.3 Leak Detection

Detection and prediction of leakages are performed through the use of the Anomaly Detection Toolkit (ADTK), a Python package for unsupervised/rule-based time series anomaly detection (ADTK, 2023). The algorithms implemented have been widely used in recent years for anomaly detection applications in time series data (Gopali and Namin, 2022), (Otte et al., 2022), (Ameli et al., 2022), (Beliichovski et al., 2022).

For all these models, there is only one hyperparameter that must be fixed that is the c-factor. The c-factor establishes when an observation has to be considered normal or anomalous basing on the historical interquartile range. We leave it to the default value c = 3.0; this means that when an observation is 3-times greater than the interquartile range of the n-previous observations, it is classified as anomalous. As Table 3 shows, the output of the trained algorithms is a data table containing for each end node, the timestamp when the leak has been identified.

Table 3: Example Output Anomaly Detection Algorithms.

End Node	Timestamp
Anfossi	2009-11-18 12:00:00
Abbiategrasso	2009-11-18 16:00:00
N05636	2009-11-19 02:00:00
...	...
N00780	2009-11-22 19:30:00

Table 4 summarizes the performances of the algorithms described in the previous section.

Table 4: Performances Anomaly Detection Algorithms on Unbalanced Dataset.

Algorithms	Accuracy	Precision	Recall	F1
PersitAD	0.8452	0.0007	0.1728	0.0015
LevelShiftAD	0.9007	0.0007	0.1043	0.0014
GeneralizedESD TestAD	0.9985	0.0004	0.0005	0.0005
InterQuartile RangeAD	0.9981	0.0029	0.0055	0.0038
Auto RegressionAD	0.8498	0.0007	0.1637	0.0018
Local Outlier Factor	0.9164	0.0005	0.068	0.0011
Isolation Forest	0.9041	0.0009	0.1325	0.0004
K-Means	0.9413	0.0001	0.0166	0.0003
Affinity Propagation	0.9911	0.0002	0.0023	0.0014

When dealing with anomaly detection problems, we would like to know how good the anomaly detection algorithm was in identifying the anomalous events. This measure is expressed by the precision.

Looking at Table 4, the InterQuartileRangeAD stands out as the top-performing algorithm in terms of precision score. However, this algorithm suffers from a Quadratic Time Complexity $O(n^2)$, requiring a greater computation effort. On the contrary, the Isolation Forest algorithm, despite its lower computational complexity (Linear Time Complexity - $O(n)$), exhibits lower precision compared to the InterQuartileRangeAD.

Difference in ranges between accuracy and the other evaluation metrics is motivated by the unbalancing of the dataset, where observations referring to normal conditions of the system are present in large quantities with respect to leakages events. The analysis of the data generated in the first tests showed that leakage duration was limited to 24 hours. It was, therefore, decided to extend this duration to obtain leakages with a minimum duration of 8 hours and a maximum of 72 hours thus increasing the size of the data set in the presence of losses.

Table 5 shows the performances of the anomaly detection algorithms tested on the balanced dataset.

Table 5: Performances Anomaly Detection Algorithms on Balanced Dataset.

Algorithms	Accuracy	Precision	Recall	F1
PersitAD	0.8450	0.4370	0.0300	0.0563
LevelShiftAD	0.8451	0.4074	0.0147	0.0284
GeneralizedESD TestAD	0.8462	0.5455	0.0040	0.0080
InterQuartile RangeAD	0.8470	0.7222	0.0087	0.0172
Auto RegressionAD	0.8403	0.3399	0.0405	0.0723
Local Outlier Factor	0.8354	0.1790	0.0194	0.0350
Isolation Forest	0.7834	0.2310	0.1752	0.1993
K-Means	0.8441	0.4383	0.0475	0.0857
Affinity Propagation	0.8443	0.3889	0.0211	0.0400

Looking at Table 5, it can be seen that the performances of all the algorithms (in terms of precision) trained on the balanced dataset improved. Also in this case, the InterQuartileRangeAD algorithm exhibits the highest precision score compared to the others. For this reason, the authors selected the InterQuartileRangeAD algorithm for the leak prediction task.

3.4 Leak Prediction

Before passing to prediction, a Feature Engineering step was needed.

Feature Engineering is a machine learning technique that leverages data to produce new information by combining features. To reach this goal, a mathematical function *f* is applied to data (Gutschi, 2018).

In this step, the authors performed a rolling window aggregation that consist in aggregating data into equally sized windows for all the timestamps. Rolling aggregation allows us to build a dataset ready for prediction so that, starting from the pressure data which represents the system behaviour over historical time widows, we can anticipate leakages occurrence.

In the analysis, the rolling aggregation process has been applied on the test dataset containing the simulated pressure data of the 24th November 2011. Table 5 reports the leak history test set.

Table 6: Leak History Test Set.

End Node	Start Time
N02197	2009-11-24 23:00:00
N15337	2009-11-24 04:30:00
N10848	2009-11-24 10:00:00
N11577	2009-11-24 16:30:00
N26758	2009-11-24 16:30:00
N06633	2009-11-24 04:00:00
N19500	2009-11-24 03:30:00
N00971	2009-11-24 03:00:00
N04628	2009-11-24 20:30:00

As Table 5 shows, the leak history dataset contains two columns: the 'End Node' that is nodes where the leak occurred, and the start time that is when the leak occurred.

For leakages prediction, two different lag windows were considered: a shorter prediction window of 1 hour and a longer prediction window with duration of 3 hours. We used the median as aggregating function since compared to the mean, it is more robust to the outliers.

4 RESULTS

In the present section we will present the results given by the application of InterQuartileRangeAD algorithm for leakages prediction.

In order to evaluate the performances of the InterQuartileRangeAD algorithm in predicting leakages, we use the start time information shown in Table 5. In other words, we will verify if the algorithm raises a warning before the true leak time.

Figure 4 and Figure 5 show the performances of InterQuantileRangeAD algorithm with 1-hour prediction window. In this case the algorithm was able to predict the leak events for two nodes of the network: N00971 and N02197. As shown in Figure 4, for the node N00971 the algorithm generated two warnings, at 02:00 pm and 02:30 pm, before the true leak time that is at 03:00 pm.

The same for the node N02197 shown in Figure 5, where the leakage event occurred at 11:00 pm and the first warning was generated by the algorithm at 10:00 pm.

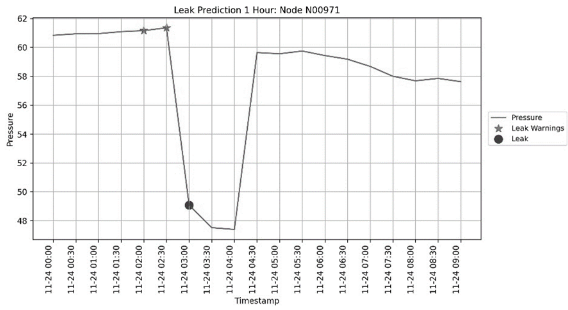

Figure 4: 1 Hour Prediction Node N00971.

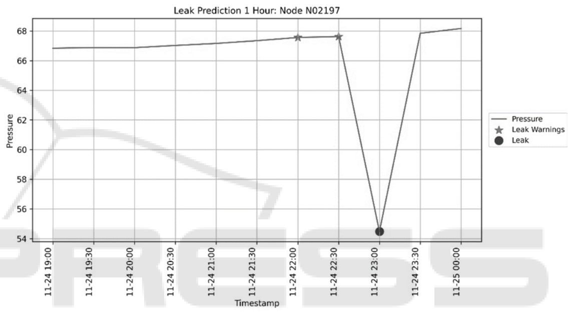

Figure 5: 1 Hour Prediction Node N02197.

In Figure 6, we report the performances of the algorithm with 3-hours prediction window. In this case, the InterQuantileRangeAD algorithm generated the first warning at midnight that is 3 hours before the true leak time (2009-11-24 03:00:00).

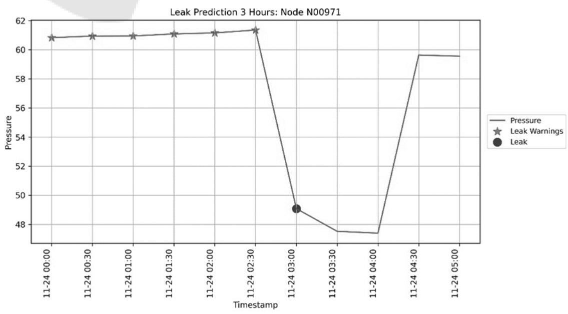

Figure 6: 3 Hours Prediction Node N00971.

5 CONCLUSIONS

In this paper an approach for leaks detection and prediction has been presented. The authors applied an unsupervised approach for detecting leakages in the Milan WDS. In our data leaks represent a minority class: as in real-world cases, values representing normal conditions of a water system are present in large quantities with respect to the ones referring to leakages, which make them an unrepresented class in data.

Unbalancing of data given as input to the trained algorithms for detection and prediction let to obtain a high gap between values of accuracy with respect to precision. To solve this problem in the present case study the leak duration has been extended, analysing leak events with a minimum duration of 8 hours. This let us have a significative improvement in the precision score of the trained algorithms. One future step could be that of considering data in the night period, normally defined as being between midnight and 6 am. During night, flowrate is low while pressure assumes maximum values. For this reason, the minimum night flow is the most meaningful piece of data as far as estimating night leakage is concerned. Giving to the algorithm flowrate data in addition to pressure data could be another possible improvement.

Finally, for future works we plan to simulate more data in order to include in the analysis weekly and yearly seasonality in water consumption.

ACKNOWLEDGEMENTS

The research results presented in this paper have been achieved inside the Water 4.0 project, named “Technologies for the convergence between industry 4.0 and the integrated water cycle”. This research project is currently running and is funded by the Ministry of Enterprises and Made in Italy (https://www.mimit.gov.it/en/).

REFERENCES

Adedeji K.B., Ponnle A. A., Abu-Mahfouz A.M., Kurien A. M. (2022). Towards Digitalization of Water Supply Systems for Sustainable Smart City Development—Water 4.0. In Applied Sciences, vol. 12, issue 18. MDPI, pp. 1-25.

Ameli M., Becker P.A., Lannkers K., Ackeren M., Bahring H., Maab W. (2022). Explainable Unsupervised Multi-Sensor Industrial Anomaly Detection and Categorization. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA).* IEEE Xplore, pp. 1468-1475.

ADTK, (2023) Anomaly Detection Toolkit 0.6.2 documentation. Available online: Anomaly Detection Toolkit (ADTK) — ADTK 0.6.2 documentation (Accessed on 28 March 2023).

Barros D., Almeida I., Zanfei A., Meirelles G., Luvizotto E. Jr., Brentan B. (2023). An Investigation on the Effect of Leakages on the Water Quality Parameters in Distribution Networks. In *Water, vol. 15, issue 324.* MDPI.

Belichovski M., Stavrov D., Donchevski S., Nadzinski G. (2022). Unsupervised Machine Learning Approach for Anomaly Detection in E-coating Plant. In *2022 IEEE 17th International Conference on Control & Automation (ICCA)*, on June 27-30 2022 (Hybrid) Naples, Italy. IEEE Xplore, pp. 992-997.

Bentoumi M., Chikouche D., Mezache A., Bakhti H. (2017). Wavelet DT method for water leak-detection using a vibration sensor: an experimental analysis. In *IET Signal Process. vol.11, issue 4.* IET Journals - The Institutions of Engineering and Technology, pp. 396-405.

Caldognetto N., Evangelisti L.P., Poltronieri F., Russo M., Stefanelli C., Tenani S., Toboli S., Tortonesi M. (2022). Water 4.0: enabling Smart Water and Environmental Data Metering. In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium.* IEEE.

Chan T. K., Chin C.S., Zhong X. (2018). Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection, In *IEEE Transactions on Knowledge and Data Engineering, vol.6.* IEEEAccess, pp. 78846-78867.

Cheng Z., Zou C., Dong J. (2019). Outlier Detection using Isolation Forest and Local Outlier Factor. In *Proceedings of International Conference on Research in Adaptive and Convergent Systems*, China, September 24–27 2019 (RACS ’19). ACM Digital Library, pp. 161-168

Cody A. R., Tolson B. A., Orchard J., Detecting Leaks in Water Distribution Pipes Using a Deep Autoencoder and Hydroacoustic Spectrograms. In *Journal of Computing in Civil Engineering, vol. 34. No. 5.* ASCE.

Cody A. R., Dey P., Narasimhan S. (2020). Linear Prediction for Leak Detection in Water Distribution Networks. In *Journal of Pipeline Systems Engineering and Practice, vol. 1, issue 1.* ASCE, pp. 1-16.

Colombo A. F., Lee P., Karney B.W. (2009). A selective literature review of transient-based leak detection methods. In *Journal of Hydro-environment Research, vol.2.* ELSEVIER, pp. 212-227.

Escofet, M.A.C., Quevedo, J., Alippi, C., Roveri, M., Puig, V., García, D., Trovò, F. (2016). Model- vs. data-based approaches applied to fault diagnosis in potable water supply networks. In *Journal of Hydroinformatics, vol.18, No.5.* IWA PUBLISHING.

Fitrianto A., Wan Muhamad W.Z.A., Kriswan S., Susetyo B. (2022). Comparing Outlier Detection Methods using Boxplot Generalized Extreme. In *Aceh International Journal of Science and Technology*, *vol.11, issue 1.* Graduate School of Syiah Kuala University, pp-38-45.

Frey B.J., Dueck D. (2007). *Clustering by Passing Messages between Data Points*. In *Science, vol. 315, issue 5814*. JSTOR, pp. 972-976.

Gansukh C., Yoo K.H, Bazarbaev M., Nasridinov A. (2021). Feasibility Study of Outlier Detection in Smart Manufacturing Applications. In *Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceeding of the 16th International Conference on IIHMSP in conjunction with the 13th international conference on FITAT*, *vol.2,* November 5-7, 2020, Vietnam. Springer, pp.283-290.

Gopali Saroj, Namin A.S. (2022). Deep Learning-Based Time-Series Analysis for Detecting Anomalies in Internet of Things. In *Electronics, vol. 11,* MDPI, pp. 1-16 .

Gutschi C., Furian N., Suschnigg J., Neubacher D, Voessner S. (2018). Log-based predictive maintenance in discrete parts manufacturing. In *Procedia CIRP, vol.79.* ELSEVIER, pp. 528-533.

Hunaidi O., Chu W.T., (1999), Acoustical characteristics of leak signals in plastic water distribution pipes. In *Applied Acoustics, vol. 58.* ELSEVIER, pp. 235-254

Kammoun M., Kammoun A., Abid M. (2022). Experiments based comparative evaluations of machine learning techniques for leak detection in water distribution systems. In *Water Supply, vol.22, Issue 1.* IWA PUBLISHING, pp. 628–642.

Kammoun M., Kammoun A., Abid M. (2022). *LSTM-AE-WLDL:* Unsupervised LSTM Auto-Encoders for Leak Detection and Location in Water Distribution Networks. In *Water Resources Management*, *vol.37, Issue 2.* Springer, pp.731-746.

Klise K., A., Murray, R., Haxton, T. (2018). An Overview of the Water Network Tool for Resilience (WNTR).

Leu S.S., Bui Q.N. (2016). Leak Prediction Model for Water Distribution Networks Created Using a Bayesian Network Learning Approach. In *Water Resources Management, vol. 30.* Springer, pp.2719-2733.

Lijuan W., Hongwei Z., Zhiguang N. (2012). Leakage Prediction Model Based on RBF Neural Network. In *Software Engineering and Knowledge Engineering: Theory and Practice, vol 114.* Springer, pp 451-458.

F.T., Ting K.M., Zhou Z.H. (2008). Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE Computer Society, pp. 413-422

Liu F.T., Ting K.M., Zhou Z.H. (2012). Isolation-Based Anomaly Detection. In *ACM Transactions on Knowledge Discovery from Data* (TKDD), *vol.6, Issue 1, No. 3*. ACM, pp.1-39.

Naiades Project. A holistic water ecosystem for digitization of urban water sector. Available online: https://www.naiades-project.eu/ (Accessed on 20 January 2023).

Otte T., Posada-Moreno A.F., Hubenthal F., Habler M., Bartels H., Abdelrazeq A., Hees F. (2022). Condition Monitoring of Rail Infrastructure and Rolling Stock using Acceleration Sensor Data of on-Rail Freight Wagons. In *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods* (*ICPRAM 2022*). SCITEPRESS, pp.432-439.

Pearson, D. (2019). Standard Definition for Water Losses, IWA Publishing, London.

Perez R., Sanz G., Puig V., Quevedo J., Escofet M..C., Nejjari F., Meseguer J., Cembrano G., Tur, J.M.M., Sarrate R. (2014). Leak Localization in Water Networks A Model-Based Methodology Using Pressure Sensors Applied to a Real Network in Barcelona. In *IEEE Control Systems Magazine, vol.34, issue 4.* Applications of Control, pp.24-36.

Philips S.J. (2002). Acceleration of K-Means and Related Clustering Algorithms. In *Algorithm Engineering and Experiments*, *4th International Workshop, ALENEX 2002*, San Francisco, CA, USA, January 4-5, 2002, Springer, pp.166-177.

Scikit-learn, Scikit-learn:machine learning in Python, 1.2.2. Available online: scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation (Accessed on 3 April 2023).

Rosner B., (1983). Percentage Points for a Generalized ESD Many Outlier Procedure. In *Technometrics, vol.25, No. 2.* ASQ, pp.165-172.

Shukla S., Naganna S. (2014). A Review on K-Means Data Clustering Approach. In *International Journal of Information & Computation Technology, vol.4, No.17,* Springer, pp.1847-1860

Tornyeviadzi H.S., Seidu R. (2023). Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data. In *Engineering Applications of Artificial Intelligence, vol.122.* ELSEVIER.

Vrachimis G.S., Kyriakou M.S., Eliades D.G.,Polycarpou M. M., (2018). LeakDB: A benchmark dataset for leakage diagnosis in water distribution networks. In *1st International WDSA / CCWI 2018 Joint Conference.*

Wan X., Kuhanestani P.K., Farmani R., Keedwell E. (2022). Literature Review of Data Analytics for Leak Detection in Water Distribution Networks: A Focus on Pressure and Flow Smart Sensors. In *Journal of Water Resources Planning and Management, vol.148, issue 10,* ASCE.

Wang L. Liu Y., Yin H., Sun W. (2022). Fault diagnosis and predictive maintenance for hydraulic system based on digital twin model. In *AIP Advances, vol. 12.* AIP Publishing.

Wang W., Sun H., Guo J., Lao L., Wu S., Zhang J. (2021). Experimental study on water pipeline leak using In-Pipe acoustic signal analysis and artificial neural network prediction. In *Measurement, vol.186.* ELSEVIER.

Xu D., Tian Y. (2015). A Comprehensive Survey of Clustering Algorithms. In *Annals of Data Science*, *vol.2, No. 2.* Springer, pp.165-193.

Yu T., Chen X., Yan W., Xu Z., Ye M. (2023). Leak detection in water distribution systems by classifying vibration signals. In *Mechanical Systems and Signal Processing, vol. 185.* ELSEVIER.