Domain-informed variational neural networks and support vector machines based leakage detection framework to augment self-healing in water distribution networks

Lauren McMillan [a,*], Jawad Fayaz [a,b], Liz Varga [a]

[a] *Infrastructure Systems Institute, Department of Civil, Environmental and Geomatic Engineering, University College London (UCL), London, UK*
[b] *School of Computing, Engineering and Digital Technologies, Teesside University (TU), Teesside, UK*

A B S T R A C T

The reduction of water leakage is essential for ensuring sustainable and resilient water supply systems. Despite recent investments in sensing technologies, pipe leakage remains a significant challenge for the water sector, particularly in developed nations like the UK, which suffer from aging water infrastructure. Conventional models and analytical methods for detecting pipe leakage often face reliability issues and are generally limited to detecting leaks during nighttime hours. Moreover, leakages are frequently detected by the customers rather than the water companies. To achieve substantial reductions in leakage and enhance public confidence in water supply and management, adopting an intelligent detection method is crucial. Such a method should effectively leverage existing sensor data for reliable leakage identification across the network. This not only helps in minimizing water loss and the associated energy costs of water treatment but also aids in steering the water sector towards a more sustainable and resilient future. As a step towards 'self-healing' water infrastructure systems, this study presents a novel framework for rapidly identifying potential leakages at the district meter area (DMA) level. The framework involves training a domain-informed variational autoencoder (VAE) for real-time dimensionality reduction of water flow time series data and developing a two-dimensional surrogate latent variable (LV) mapping which sufficiently and efficiently captures the distinct characteristics of leakage and regular (non-leakage) flow. The domain-informed training employs a novel loss function that ensures a distinct but regulated LV space for the two classes of flow groupings (i.e., leakage and non-leakage). Subsequently, a binary SVM classifier is used to provide a hyperplane for separating the two classes of LVs corresponding to the flow groupings. Hence, the proposed framework can be efficiently utilised to classify the incoming flow as leakage or non-leakage based on the encoded surrogates LVs of the flow time series using the trained VAE encoder. The framework is trained and tested on a dataset of over 2000 DMAs in North Yorkshire, UK, containing water flow time series recorded at 15-minute intervals over one year. The framework performs exceptionally well for both regular and leakage water flow groupings with a classification accuracy of over 98 % on the unobserved test dataset

1. Introduction

Water distribution networks face growing pressures from rising population levels, increased urbanisation, more significant uncertainty in supply due to climate change, rising energy prices, a volatile global economy, and a more complex social and regulatory environment (Marlow et al., 2013). These stressors have seen the industry shift from a 'predict and provide' approach, which does not account for limitations in supply, towards a 'conserve and contain' approach (Taylor and Hodges, 2008). This change in mindset seeks to reduce per capita consumption through a combination of demand-side behaviour change and improved supply-side efficiency. Leakage due to pipe bursts is a major inefficiency in water distribution networks and over 20 % of the water entering public supply in England and Wales is lost as leakage, a wastage of over 50 litres per person per day (PR19 final determinations 2019).

A priority for Ofwat, the economic regulator of the water industry in England and Wales, is to reduce leakage across water distribution networks. High levels of leakage can undermine efforts to reduce consumption on the demand side, as consumer confidence in their water

Nomenclature

AE	autoencoder
ANN	artificial neural network
AUC	area under the curve
DMA	district meter area
IA	index of agreement
KF	Kalman filter
KL	Kullback–Leibler
LKG	leakage grouping
LV	latent variable
MNF	minimum night flow
NLKG	non-leakage grouping
PDF	probability density function
ROC	receiver operating characteristic
SVM	support vector machine
VAE	variational autoencoder

supplier is diminished. This is compounded by the fact that, despite efforts by water companies to improve leakage management, a significant proportion of leakage bursts in distribution systems are reported by consumers rather than detected by the companies themselves. While catastrophic failures can have an average response time of 1.5 to 3.5 hours, depending on the time of day, bursts that are not visible from the surface may go undetected for months in the absence of customer reporting (Mounce et al., 2010). Traditional leakage identification and analysis methods are unable to offer a rapid response, typically relying on nighttime sensor data to spot leakage. New sensing technologies, while promising, will take decades to see widespread distribution across an extensive buried network. Hence, more proactive and reliable approaches are needed for leakage management in water distribution systems, that are able to make use of existing sensor technologies ("'Proactive' approach to leaks required to meet tough Ofwat targets 2023).

The standard practice for water utility companies in the UK is to divide the water distribution network into district meter areas (DMAs). DMAs represent isolated water network areas, typically serving up to 2000 households, where sensors at the inlet and outlet measure the flow. Leakage management is usually performed at the DMA level (Morrison, 2004). In the field of leakage management of water flow distribution networks, leakage detection is a critical research subject (Mounce et al., 2010; Romano et al., 2014; Chan et al., 2018; Puust et al., 2010; Aksela et al., 2009). Sensor data can be fed into leakage detection models that seek to identify bursts by monitoring changes in the flow profile over a set window of time. Traditionally, the most common methods for identifying leaks utilise minimum night flow (MNF) (García et al., 2008). This technique recognises that water usage during night-time is less variable than in the daytime. Hence, the average nightly minimum over a specified window is used as a baseline for comparison with new flow data, with a significant variation (relative to a pre-defined threshold) indicating a leak (Mounce et al., 2007). However, these techniques are not highly reliable as MNF methodologies have to deal with several uncertainties (such as seasons, weather conditions, holiday periods, gatherings, etc). Accurate use of MNF relies upon having sufficient knowledge to estimate several parameters, including active night users, leakage exponent (which varies with system pressure), and the hour-to-day factor (Amoatey et al., 2021). Reliable estimation of these parameters typically requires both pressure and flow data. The selection of the best time window for the computation of MNF requires additional considerations and analysis. It has been shown that minimum error does not correspond with the selected night flow window but with the hour in which average demand applies (García et al., 2008). While it is often the responsibility of trained operators to identify leakage from MNF, a

significant proportion of leaks are reported to water companies by their customers (Mounce et al., 2007).

MNF is a leakage detection method based on simplified demand forecasting, as the average of the time window is assumed to be the required demand for future nights. Deviation from expected demand is a method of leakage identification. Demand forecasting therefore has a potential role to play in developing leakage detection methods that are based on comparison between expected and actual flow. Methods for demand forecasting in water distribution networks include both traditional and machine learning-based forecasting techniques (Antunes et al., 2018). Hybrid models have also been developed which incorporate machine learning methods, and these have the potential to further improve forecasting accuracy (Pandey et al., 2021; Pu et al., 2023).

Recent works have proposed models based on machine-learning methods to improve the accuracy and reliability of leakage detection (Pu et al., 2023; Abdelmageed et al., 2022; Fu et al., 2022). Benefits of intelligent methods include increased automation, high levels of accuracy and reliability, ability to deal with high temporal resolution, and more rapid leakage identification (Fu et al., 2022). Some of the machine-learning and deep-learning techniques utilised by these studies include artificial neural networks (ANNs) (Mounce and Machell, 2006; Aksela et al., 2009; Romano et al., 2014; Zhou et al., 2019), support vector machines (SVM) (Geberemariam et al., 2014; Kang et al., 2018), Kalman filters (KFs) (Ye and Fenner, Feb. 2011; Jung and Lansey, 2015), and wavelet analysis (Romano et al., 2014). With sufficient quality and quantity of training data, these methods have demonstrated strong performance in leakage identification (Mounce and Machell, 2006). Many of these models are trained using examples of standard flow data and flow during leakage bursts. The burst examples are typically obtained by matching the timestamps of abnormal flow patterns to pipe repair records or reports of visible leakage from consumers to water companies. Alternatively, the data can be simulated through a hydrant flush event that mimics a leakage burst (Birek et al., 2014). Some studies do not use data from real water distribution networks and instead extract pressure data from simulation software-based network models (Leu-and Bui. 2016).

In the field of deep-learning, autoencoders (AE) are relatively novel and draw upon the concept of dimensionality reduction using artificial neural networks (ANNs) with bottleneck shapes at the central layers of the ANN (Hinton and Salakhutdinov, 2006). Variational autoencoders (VAE) are a type of AE that relies on Bayesian concepts and forces the bottleneck layers to possess a regularised standard normal space (Kingma and Welling, 2019). This reduces the dimensions of input data in such a way that the inputs similar to each other in terms of their characteristics, lead to similarities in the outputs of the bottleneck layers (Kingma and Welling, 2019). Hence, within the setting of water leakage detection, it can be understood that with sufficient training of a VAE using leakage and non-leakage flow datasets, VAEs can be capable of differentiating between the flow data classified as leakage or non-leakage. VAEs have demonstrated their potential in the detection of extreme events in numerous engineering contexts (González-Muñiz et al., 2022), including earthquake early-warning systems (Fayaz and Galasso, 2023), detection of cyber-attacks (Zavrak and İskefiyeli, 2020), and structural health monitoring of infrastructure such as dams (Shu et al., 2023). Various types of AE, including VAEs, have begun to be considered a tool for leakage detection in both water and oil/gas pipelines, where they have shown initial promise (Cody et al., 2020, Wang et al., 2020). However, previous studies using AE for leakage detection have relied on test-bed setups, where water flow behaviour can be strictly controlled. These setups vary in scale, from representing a single component (Feng et al., 2021, Ahmad et al., 2023), a handful of pipes (Yeo et al., 2019), to a broader distribution network more comparable in scale to a small DMA (Cody et al., 2020). Some setups model only regular flow (Yeo et al., 2019), while others simulate leakage events (Cody et al., 2020, Feng et al., 2021). Such setups allow for cutting-edge sensor technology to be used, and so all work to date has used hydroacoustic

measurements from acoustic sensors (Cody et al., 2020, Feng et al., 2021, Ahmad et al., 2023, Yeo et al., 2019) rather than traditional water flow measurements. While this can provide more accurate sensing, acoustic sensing is a less explored method of monitoring behaviour in water pipes. The technology has undergone limited deployment, and most DMAs are not subjected to acoustic monitoring. In developed nations with well-established water distribution systems, the cost of installing improved sensing technology across the entire existing network is high – one supplier for North West England spent £30 million installing 100,000 acoustic loggers across its network (United Utilities - World's biggest listening project helps tackle water leaks, 2019) – and water companies report large numbers of their acoustic loggers not working due to failed batteries, incorrect attachment, communication failure, etc. (Pressure logging or acoustic logging, 2020).

Hence, this study explores machine learning and deep learning methods for widespread water flow monitoring at a DMA level. In particular, this study proposes a VAE-based framework that reduces the dimensions of incoming real-time flow time series into two surrogate latent variables (LV) space. This is done by training a VAE using ~10,000 groupings of 96-point flow time series from ~2000 DMAs in the North Yorkshire region of England, UK, and utilising a novel domain-informed loss function. The loss function is tuned based on the application-specific assumption that the flow during leakages possesses different characteristics than the usual/normal flow. Thus, a new loss term is introduced in the original VAE loss function to penalise any overlap of the LVs corresponding to the leakage (LKG) and non-leakage (NLKG) flow groups. Hence, the trained surrogate LV space possesses distinct yet regulated LVs for the two types of flow groupings. The encoded LV space is then used to train a binary SVM classifier which can be used to accurately classify the incoming encoded LVs into LKG or NLKG in real time based on its mapping onto the pre-trained LV space. Unlike previous studies using test-bed setups, this study uses real-world DMA flow data. This could yield the environmental and economic benefits of reduced water loss (and the associated energy and resource savings) without requiring the cost of widespread deployment of new sensing technologies. Hence, the proposed framework can provide a greater level of resilience in existing and aged water infrastructure systems where uptake of new sensing technologies is likely to be gradual. It is verified that the proposed framework is highly successful in accurately monitoring the leakage status of DMAs and can handle the additional challenges of processing noisy real-world data.

2. Framework conceptualisation

In complex infrastructure systems, a vast number of individual components are often difficult to access (e.g., buried infrastructure). Hence, directly detecting failure via inspection can be prohibitively expensive and, to some extent, relies on noticeable/surface-level defects within the system. This can be costly to the resilience of both the system and the societies it serves and result in unsustainable and non-climate-friendly wastages, as well as monetary losses. Therefore, instead of directly observing failure in water infrastructural systems, operators rely on the data, often in the form of time series data, from a sensor network to try and identify failure events. In such cases, data-driven- and machine learning-based models can offer a robust solution to the problem of failure detection (Chan et al., 2018; Puust et al., 2010).

An issue with time series data can be the curse of high dimensionality (Bellman, 1961) and the difficulty of developing damage detection cut-offs. Due to this, differentiating between critical information and noise in the data becomes challenging. To address this issue, many machine learning algorithms require some feature engineering, which is the application of domain knowledge to identify and select a subset of features from a data set (e.g., mean, variance) to be used as inputs to train these algorithms. In other cases, algorithms like principal component analysis (PCA), singular value decomposition (SVD), etc. conducts feature reduction by statistically merging original features into lower

dimensions. Within this setting, AE neural networks (particularly VAEs) conduct effective dimensionality reduction without requiring explicit feature engineering. In addressing the challenges posed by time series data, selecting VAEs over other methods, such as wavelet transform andfunctional principal component analysis (FPCA), offers several distinct advantages (Todo et al., 2022). Time series data often exhibits complex and non-linear relationships, and VAEs can capture these non-linear dependencies through their neural network architectures, allowing them to represent the data in a more flexible manner and with minimal assumptions. Furthermore, VAEs learn a probabilistic latent space representation of the data, providing a distribution of possible representations for each input, which is valuable when dealing with noisy or imperfect sensor data. Specifically, VAEs can produce smooth and regulated LVs that act as statistical surrogates for the input data by reducing the dimensions and capturing the key characteristics. Noteworthy alternatives like PCA, SVD, and regular AEs, though valid, may not inherently provide this continuous probabilistic dimensionality reduction with comparable precision and reconstruction efficiency. Hence training and exploring the LV space of the VAEs can provide high-dimensional insights within a low-dimensional space where the distance between the LVs indicates the similarity/dissimilarity within the characteristics of the input data. This offers an approach to identify any anomalies/failures (such as pipe bursts) in a lower-dimensional surrogate space rather than the original complex and high-dimensional space of the inputs (especially water flow in water infrastructure systems).

With this backdrop, this study proposes a framework for leakage identification based on the concept of statistical surrogacy. Rather than directly classifying the original high-dimensional water flow time series data into LKG or NLKG categories, the framework instead reduces the dimensions of the flow data using a domain-informed VAE to minimise the impact of redundant characteristics and isolate the key components of different classes through surrogate LVs. Classification can then be performed on the surrogate LVs, which are trained to capture the distinction between the LKG and NLKG flow groupings through a domain-informed loss function. VAE is chosen for obtaining the surrogate variables as they have proved to be effective method for natural data (van der Maaten et al., 2009) and in cases of extreme events (Fayaz and Galasso, 2023), with leakage representing an extreme case of flow behaviour. Furthermore, due to the Bayesian nature of the VAE, the flexibility of altering the loss function provides an efficient solution to include the physics of the problem in the training process.

The proposed framework is illustrated in Fig. 1. Sensors record the net flow of water for a given DMA at a discrete time interval (e.g., every 15 min in this case). The proposed framework uses a pre-specified length of the water flow record (e.g., preceding 24 h, i.e., 96 points, in this case) and classifies them into LKG or NLKG flow in real time using end-to-end pre-trained models of VAE and SVM. The framework starts by converting the preceding flow time series data into two surrogate mean LVs (i.e., μ_{LV_1} and μ_{LV_2}) using a pre-trained VAE encoder. The LVs are trained to be
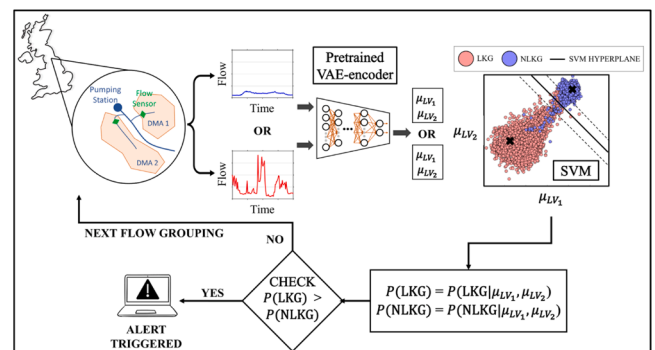

Fig. 1. The proposed VAE and SVM-based framework.

sufficient and efficient to contain information about the required characteristics of the water flow. The obtained μ_{LV_1} and μ_{LV_2} are then used as inputs to a pre-trained SVM, which compares it against the pre-obtained mapping of LVs to compute the probability of the flow data to be classified as LKG ('burst') or NLKG ('usual') flow (i.e., $P(\text{LKG}) = P(\text{LKG}|\mu_{LV_1} \text{ and } \mu_{LV_2})$ and $P(\text{NLKG}) = P(\text{NLKG}|\mu_{LV_1} \text{ and } \mu_{LV_2})$). Then $P(\text{LKG})$ is compared against $P(\text{NLKG})$, and based on this comparison, the final classification decision is made (i.e., LKG/burst or NLKG/usual flow). Thus, this framework enables rapid monitoring of the water pipes and flags possible leakages without human intervention. This information can then be used to inform the targeted repair strategies that minimise water loss in the network and curtail inconvenience to the public.

3. Dataset

This study uses a water flow dataset of ~2500 DMAs managed by Yorkshire Water, a utility company of water supply and distribution in North Yorkshire, UK. The dataset consists of water net flow (in litres/second) for each DMA recorded at a 15 min interval from April 2016 to April 2017 (~ 365 days × 24 h × 60 min/15 min interval = 35,040 data points). Each data point is given a validity code based on Yorkshire Water's assessment of the sensors' records. These codes – 'V' for valid, 'I' for invalid, or 'M' for missing – reflect any possible breaks or faults in the sensor readings. Invalid or missing sections represent less than 5 % of all DMA flow data. Fig. 2a shows a full year of flow data for one exemplar DMA. The magnitude of flow remains broadly consistent throughout the year, with some seasonal fluctuations and spikes of large flow rates. Fig. 2b shows a standard week of valid flow data from the same DMA. The figure shows the typical volatility over a 24 h period of flow data; minimums are seen during the night hours, with peaks occurring during the morning and late afternoon that correspond with a large proportion of the population leaving for and returning from work/school.

The provided dataset also contains the repair logs during the one-year period with corresponding repair dates for respective DMAs (although the exact timestamps of repairs are unknown). The logs contain the dates of over 5000 recorded repairs across 1600+ unique DMAs. Although this repair log does not explain each repair's reasons, it is assumed that the entries are mainly due to leakage/burst events. Repairs are typically prompted either by customer leakage reports or the identification of unusual flow data by Yorkshire Water operators. If a leak is customer-reported and visible at the surface level, it is often repaired within a few hours or days. However, no evident leakages may take several weeks to repair. This delay between the onset of the leakage/burst event and the repair date means that a direct comparison of flow and repair logs is insufficient to tag the leakage dates in the dataset. Instead, it is necessary to utilise a method that identifies "abnormal" flow data representing probable leakages/bursts. The

timing of this flow can then be compared to recorded repair logs to ensure that the identified bursts are within the vicinity of the closest logged repair date (after the burst). In the absence of widespread metering, repair logs are the best alternative for the verification of identifying leakage events.

4. Dataset pre-processing

In an ideal scenerio, the framework would be trained on a dataset of confirmed leakages (as well as confirmed non-leakage flow datapoints) drawn from a complete dataset without any missing or invalid water flow data. However, such an ideal dataset is unrealistic due to the various aberrations and data errors that occur in typical real-world raw data. Hence, it is necessary to statistically complete the available water flow data and to appropriately sample examples of water flow data to represent bursts/LKG| events and periods of regular/NLKG flow. This section outlines the pre-processing required to generate the data inputs necessary to train the proposed framework.

4.1. Data completion using Kalman smoothing

As mentioned in the previous section, the raw sensor flow data provided by Yorkshire Water contains faulty segments labelled as "invalid" or "missing" or containing impossible flow values (such as negative flow). As this study proposes a data-driven framework, ensuring the dataset is robust and doesn't contain any invalid/missing data is vital. Hence, before utilising the dataset to develop the framework, the fault segments of the flow time series are corrected using Kalman smoothing (Shumway and Stoffer, 1982).

Kalman smoothing is based on the technique of KF, a simple dynamic Bayesian network that uses observed measurements (assumed to be a combination of state and noise) to provide recursive estimates of the underlying state at each time-step t (Masreliez and Martin, 1977). The KF process consists: i) a prediction step, to estimate the underlying state and covariance, and ii) an update step, which uses information from the observed measurement (at time-step t) to revise these estimates (Durbin and Koopman, 2012). Eqs. (1) and 2 are used in KF to represent the observation and the state of time series data, where X_t is the observed (or measured) value at time-step t, y_t represents the underlying state, θ is a tuning parameter, and v_t and w_t are noise components that are assumed to be normally distributed with a mean of 0 and standard deviations of Φ and τ, respectively.

$$X_t = y_t + v_t \left(\sim N\left(0, \phi^2\right) \right) \tag{1}$$

$$y_t = \theta y_{t-1} + w_t \left(\sim N\left(0, \tau^2\right) \right) \tag{2}$$

Kalman smoothing is a post-processing method that estimates the state of time series data before and after a given smoothing window and
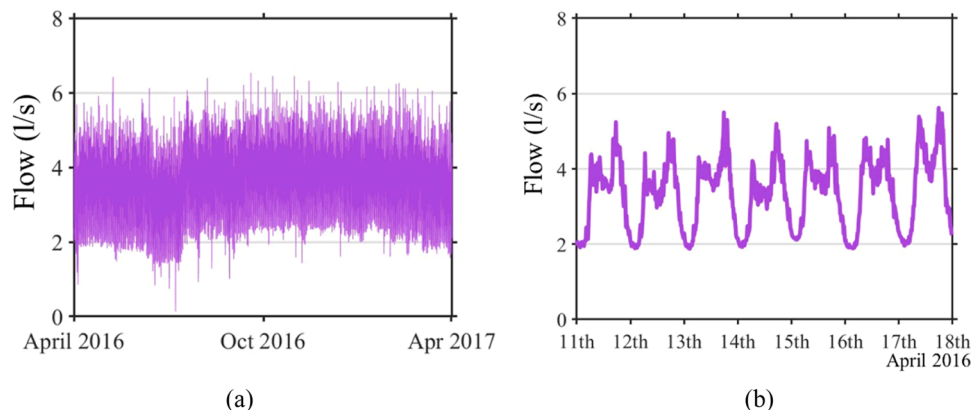

Fig. 2. Flow data from DMA 5 for (a) a full year and (b) a typical week.

performs Bayesian-state interpolation of the observations. For a given window ($t = 1, 2, \ldots, T$), a forward pass of the time series is completed with KF, followed by a backward recursive pass. This backward pass allows estimates to be refined using information from later observations after the smoothing window ($t > T$) (Briers et al., 2009).

Kalman smoothing is utilised to replace all the faulty segments in the datasets. An initial examination of the dataset used in this study reveals that approximately 95 % of segments with missing or incorrect flow data consist of fewer than 480 data points, which is equivalent to a span of 5 days when compared to the total available water flow data spanning a year. Moreover, more than 85 % of these segments contain less than 96 data points, corresponding to a duration of 24 h. The median and mode values for the missing or invalid segments are both observed to be 5 data points, equivalent to 75 min. Consequently, the presence of such a small proportion of missing or invalid data is not expected to adversely impact the Kalman smoothening process. It is also worth noting that, of the complete set of available flow data from the ~2000 DMAs, only ~10,000 groupings representing both LKG and NLKG flow patterns are employed to train the framework (explained in section 4.3). Therefore, the influence of missing or invalid data is further constrained.

Fig. 3 shows a short segment with missing flow data from an exemplar DMA (a) before and (b) after Kalman smoothing. As seen in Fig. 3a, the flow data contains a missing section, which is completed with a smooth flow profile that connects the preceding and subsequent data using Kalman smoothing in Fig. 3b. Thus, Kalman smoothing can replace erroneous sensor data with realistic values based on the available non-erroneous data. This process is repeated for all faulty segments of the dataset individually.

4.2. Outlier labelling using isolation forest

Ideally, the flow data during known leakage events would be flagged as LKG in the dataset. However, bursts in real DMA networks are rarely so neatly catalogued, with most leakage events being identified in the aftermath through customer reporting. Therefore, the best available verification for the known leakages in the available dataset is assumed to be the recorded repair log. However, the repair logs do not neccesarily correspond to the leakage timestamps (rather only contain repair dates), and the actual timestamps of leakages are still unknown. Hence, it becomes necessary to use post-hoc anomaly/outlier detection algorithms to statistically label the most probable timestamps of leakage (Mounce et al., 2007, Mounce et al., 2013). In this study, continous sets of statistical outliers are treated as leakage events and the ourliers are identified using a tree-based unsupervised machine learning algorithm: isolation forest (Liu et al., 2008). Isolation forest assumes that outliers will be rarer than the expected datapoints and have different attributes, making the outliers easier to isolate. In terms of decision trees, this phenomenom leads to outliers being placed closer to the root node than the normal data points. The classification threshold, which separates

outliers from non-outliers, is set by a hyperparameter called contamination fraction (Liu et al., 2012). In this study, a contamination fraction of 0.005 is selected. This algorithm is used to label the flow data for all DMAs as outliers and non-outliers, and the outliers are further analysed to validate the indication of potential leakages.

Fig. 4 presents the outliers detected in the flow data of one of the most-repaired DMAs. The dashed lines in the figure show the dates of repair based on the repair log, while the green circles are the outliers flagged by the isolation forest. It can be observed that the isolation forest algorithm performs well in identifying both extreme outliers and extended periods of unusual flow rates. The detected outliers, particularly extreme ones, correlate well with repair dates. While a few repair dates are observed to be away from the outlier data, this can be due to the repairs being conducted for reasons other than pipe leakage, such as replacing aging infrastructure or capacity upgrades, which are not of interest in this study. It is further observed that the algorithm also flags some other unusual flow data points that do not appear to be leakages as they aren't close to the repair dates. Hence, it is crucial to identify outlier groups so that only extended periods of irregular flow are flagged as outliers hence leakages, while isolated individual outliers are discarded. For this reason, leakage groupings labelled LKG are required to have a minimum of 20 outliers in length, representing five hours of flow. The literature supports this approach, suggesting that abnormal flow shorter than a few hours in size is not likelyto indicate leakage but sensor error, firefighting, or an industrial event (Mounce et al., 2007).

Identifying accurate leakage data points is essential in developing a reliable tool for classifying LKG and NLKG data. To validate the assumption that the detected outliers can act as a reliable proxy for the
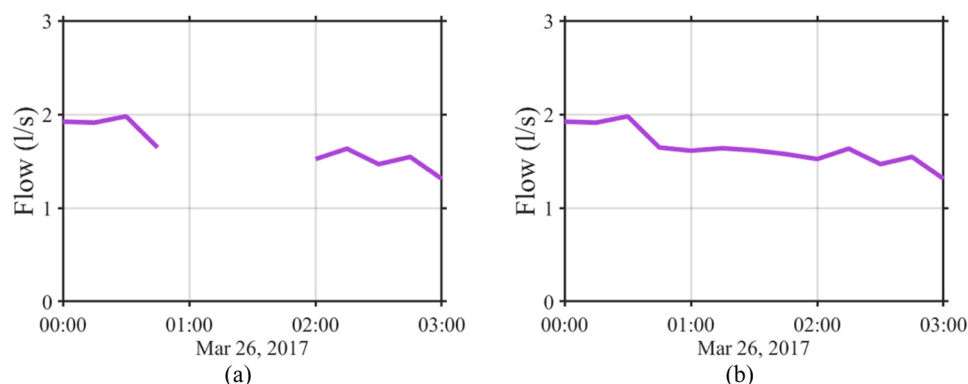

Fig. 4. Outlier identification and repair dates for DMA 2131.

Fig. 3. Section of missing data (a) before and (b) after Kalman smoothing, DMA 1348.

true leakage events, the timesteps at which the outliers are flagged using the algorithm are compared to the repair dates in the repair log. Though these timings are not expected to align perfectly due to fluctuations in the time taken to respond to suspected leakage, a reasonable time frame is necessary (a time frame of 30 days is used here). Furthermore, it should be noted that the repair logs only contain the repair dates (rather than exact timestamps); hence time lags are expected.

Fig. 5 shows the time difference between each repair date (assumed to be 23:59:59 h of each repair date) and the closest outliers before and after the repair dates for DMA 2131. The DMA has undergone 17 repairs (shown on the *y*-axis) during the year. As the repair logs contain only the date of repair, and not the time, this analysis assumed that each repair occurred at 23:59:59 h; hence the outliers occurring on the same date as a repair are recorded to occur prior to the repair. Based on Fig. 5, it is noted that the outliers correspond well with documented repairs, with many repairs occurring within two to three days of an outlier, which are likely to represent repairs to customer-reported, surface-visible bursts. Almost all recorded repairs in the example shown in Fig. 5 took place less than five days after a record of outlier flow. This falls well within the repair timescale that would be expected for less urgent, non-visible leakage or leakage on land requiring permissions for access. These findings confirm that repair data is the best proxy for leakages/bursts.

4.3. Selection of training and testing datasets

Once the outliers have been identified and examined, they must be grouped into LKG/burst and NLKG/usual groupings to provide training samples for the VAE and SVM. The outliers close to each other in time (within a few hours of each other) are likely to indicate a single burst rather than more distinct bursts and hence are grouped together. Single outlier points may indicate sensor error or data quality issues. In addition, the literature indicates that short periods of anomalous flow, lasting just a few hours, can often be attributed to industrial or fire-fighting events rather than leakage (Mounce et al., 2007). Hence, a minimum length of outlier grouping is required to ensure outlier groupings likely represent leakage. For this study, a minimum length of five hours of flow data with outliers is qualified as LKG groupings. Based on this criterion, ~3500 LKG groupings are identified for all DMAs combined, ranging from five hours to ~3.5 days of outliers. The choice of a five-hour duration for leakage groupings aligns with both literature insights (Mounce et al., 2007) and expert consultations. Notably, the proposed framework's flexibility allows for adjustments to shorter leakage groupings if additional data were accessible to guide their determination. In instances where brief yet substantial leakages are officially reported and verified by water utilities, the associated flow data could be employed as training examples. This is particularly significant as, in a real-world network deployment, the framework's adaptability could evolve over time. After the detection and repair of a leakage, coupled with the recording of its corresponding flow data, this
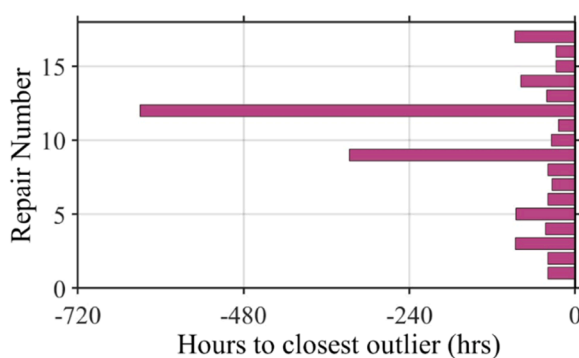
information could be utilised to retrain the framework. Consequently, this approach has the potential to enhance the accuracy of identifying similar events in subsequent occurrences.

Fig. 6 shows the distribution in the length of the LKG groupings. It is observed that outliers have a large range of lengths, though shorter outliers of five to ten hours are far more common. Indeed, over 80 % of outlier groupings contain less than 12 h, or 48 points, of flow data, and over 90 % of outlier groupings contain less than 24 h, or 96 points, of flow data.

In order to train the VAE, all input sequences must have the same length. Hence, LKG groupings are padded with zeros (after the flow data) to make them up to the length of the largest LKG grouping. Finally, in order to limit the impact of the zero-padding on the training of the VAE, a maximum limit of LKG grouping length is set. Various length cutoffs were examined, including 48, 64, 96, 128, 160, and 192 datapoints, in order to identify the optimal trade-off between early warning time, classification accuracy, and effective dimensionality reduction. The investigation revealed that a length cutoff of 96 datapoints yielded the most favourable equilibrium among these considerations. Accordingly, this specific cutoff value, equivalent to a span of 24 h of flow data, was selected as the preferred choice for this study. It is evident from Fig. 6 that over 90 % of LKG groupings contained less than 24 h of data, and hence this cutoff achieves the compromise of retaining a representative sample of LKG groupings while restricting the impact of zero-padding on the training of the VAE. The limited groupings over 96 datapoints were discarded to ensure that the leakage characteristics are fully captured within the selected cutoff. Thus, around 3500 potential LKG groupings are identified.

As the proposed framework aims to classify both regular flow data and leakage flow data, ~7000 NLKG groupings (twice the number of LKG groupings) with length equal to the LKG groupings (i.e., 24 h of flow, 96 points) are randomly selected and combined with LKG data to train the VAE. To ensure that the input data does not contain any erroneous data points (possibly due to undetected sensor error), variance checks are performed on the NLKG groupings, and those with coefficients of variation (COV) greater than 100 or less than 0.01 are removed. All these checks finally lead to a final dataset of 3336 LKG groupings and 6818 NLKG groupings. This results in a total of ~10,000 flow times series LKG and NLKG groupings for training the proposed framework. Fig. 7 shows an example of an (a) LKG and (b) NLKG grouping, for input into the VAE. It can be observed that the LKG example has significantly more variability (with flow values ranging from below 10 l/s to over 80 l/s), while the NLKG example exhibits much less fluctuation (remaining below 10 l/s throughout the 24 h period). The peak flow of the LKG example is also over eight times greater than the peak flow of the NLKG example. This contrast in flow behaviour is common between LKG and NLKG groupings, with LKG
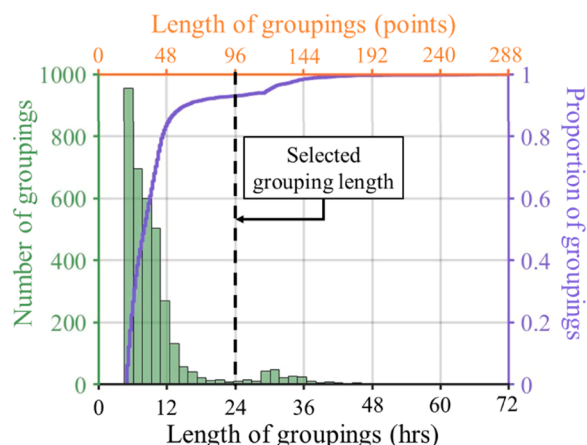

Fig. 5. Time difference between repairs and closest outliers (hrs) for DMA 2131.

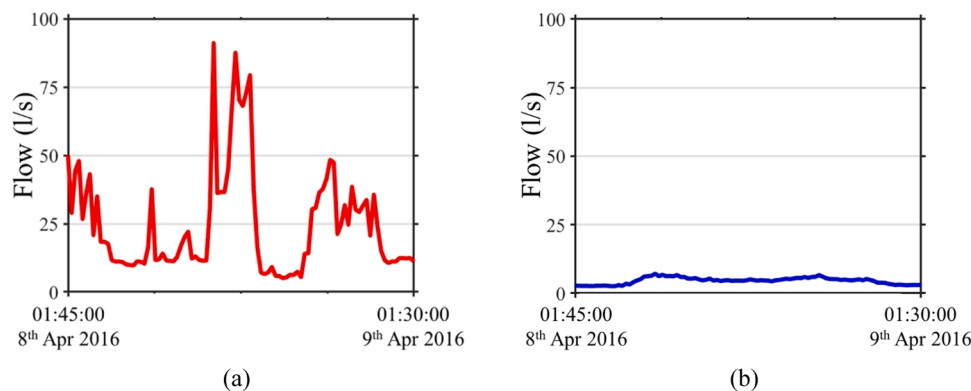**Fig. 6.** Length (in hours and number of data points) of LKG groupings.

Fig. 7. Examples of (a) LKG and (b) NLKG groupings for training the proposed framework.

groupings typically having higher maximum flow values and greater variability in flow values.

5. Training of the framework

This section outlines the general procedure for training the proposed VAE-SVM framework, with the details of VAE and SVM explained in the respective sub-sections. As discussed in Section 4, the ~10,000 LKG and NLKG flow groupings from over 2000 DMAs are carefully processed and selected to train the proposed framework. The flow time-series groupings are randomly split into train and test datasets. The train dataset is used to train the VAE-SVM framework. VAE aims to reduce the dimensionality of 96×1 incoming flow time series data into two sufficient and efficient surrogate LVs. The LVs map the flow time-series onto a regularised two-dimensional variable space such that the LVs of LKG and NLKG groupings are maximally separated using a domain informed loss function. The relative position of the LVs is based on the similarity/ dissimilarity of the time series groupings, which can be easily used to deduce the type of flow grouping. Hence, the trained VAE projects the flow time series to a two-LV space. Then, an SVM classifier is used to develop a decision boundary between the LVs of LKG and NLKG flow groupings to classify the LV into bursts or usual flow. Once trained, the framework can map unlabelled flow time series groupings onto the LV space and then probabilistically classify them as LKG or NLKG based on their position relative to the decision boundary. The framework's two principal components, i.e., VAE and SVM, are described in the following sections.

5.1. Variational autoencoder (VAE)

VAEs are from the family of Bayesian neural networks, and their premise is based on AE neural networks (Kingma and Welling, 2019). AEs are a type of neural network used for the dimensionality reduction of vectorial data and are often used to find efficient data representations (Hinton and Salakhutdinov, Jul. 2006). Due to high temporal dimensions of the flow data (curse of dimensionality), dimensionality reduction plays a crucial role in the proposed analysis. Training a classification model directly on the time series data without dimensionality reduction would be computationally inefficient (Verleysen and François, 2005). This challenge stems from the exponential increase in data sparsity and computational complexity as the number of dimensions increases, which severely impacts the model's computational feasibility and performance. Preliminary analyses using an SVM and logistic regression classifiers on the original 96-point flow data underscore a substantial bias intrinsic to the algorithms' handling of a high-dimensional feature space. This bias culminates in a pronounced underfitting tendency, where the two algorithms struggle to accommodate the vast number of dimensions adequately. The bias-related underfitting can be ascribed to the fact that these machine learnin

algorithms weigh all 96 individual data points uniformly, irrespective of their varying levels of significance in capturing leakage patterns. Crucially, not all 96 data points equally manifest pronounced leakage trends; instead, leakage patterns collectively emerge through intricate interactions across the 96-dimensional space, each with differing degrees of deviation from established norms. This intricate interplay poses a challenge when the SVM and logistic regression are directly applied to the original data, as the classifiers lack the discriminatory power to segregate the informative leakage-influenced data points from the broader set of 96 points. This limitation is particularly evident in the context of subtle leakage patterns that may span multiple data points and dimensions.

In essence, a standalone machine learning based classifier struggles to distinguish the fine-grained leakage characteristics within the original 96-dimensional space, inhibiting its capacity to accurately discern and classify nuanced leakage events. Thus, dimensionality reduction emerges as a pivotal strategy to address these limitations and bolster the efficacy of the leakage detection framework. By reducing the dimensionality of the data while retaining its essential characteristics, dimensionality reduction techniques such as VAEs enable a more concentrated representation that facilitates the identification of key patterns related to leakages. This transformed space retains the crucial leakage-related information while mitigating the curse of dimensionality,

AEs consist of a neural network-based encoder trained with a neural network-based decoder. The encoder reduces the dimensionality of vectorial input data to produce LVs, a lower-dimensional embedding that seeks to capture the defining characteristics of the input data. The choice of LV dimensions is made based on the trade-off between the reconstruction power and explainability/visualisation of the LVs. Hence, this study uses a two-dimensional LV space to provide sufficient reconstruction power while ensuring the results are interpretable and explainable. The decoder then uses the LV space to reconstruct the input data effectively with minimal loss. While a standard AE maps the input data onto a deterministic LV space, in a VAE (Kingma and Welling, 2019), the input data is instead mapped onto a probabilistic LV space with a pre-defined probability distribution and the LV space is compelled to possess smooth and continuous representations.

Consequently, points in closer proximity in the latent space lead to similar reconstructions using the decoder. This is done using a neural network-based encoder (recognition model) trained with a neural network-based decoder (generative model) that can use the LV space to reconstruct the observations. This means that the encoder describes a probability distribution for each latent attribute from which values are randomly sampled to be fed into the decoder that is expected to accurately reconstruct the input. The LVs space is constructed using Bayes' rule given by Eq. (3), where X represents the input vector (in this case 96×1 flow groupings).

$$p(LVs|X) = \frac{p(X|LVs)p(LVs)}{p(X)} \qquad (3)$$

Traditionally, VAEs are trained using a loss function consisting of two terms: i) reconstruction loss (denoted as $Loss_{recon}$) and ii) the Kullback–Leibler (KL) divergence loss (denoted as KL_{LV}) (Kullback and Leibler, 1951, Asperti and Trentin, 2020). $Loss_{recon}$ is the average of the mean squared error across the input and output (reconstructed input) vectors and measures how accurately the network reconstructs the original data (expressed in Eq. (4) where n is the total number of input sequences, i is the sequence of interest, and X and $\widehat{X}$ are the true and reconstructed vectors of time series data respectively). On the other hand, KL_{LV} measures how closely the generated LVs match the target probability distribution (typically standard normal distribution as expressed in Eq. (5) where n is the total number of input sequences, i is the sequence of interest, and μ and σ are the mean and standard deviation vectors of the generated LVs, respectively). KL divergence is a directed distance measure that determines the deviation of one probability distribution compared to the other. Therefore, the higher the KL divergence, the higher the deviation between the two distributions. In other words, $Loss_{recon}$ makes sure that the LVs are sufficient and efficient representations of the input data X while KL_{LV} forces the LVs to possess a smooth and regularised target distribution space.

$$Loss_{recon} = \frac{1}{n} \sum_{i}^{n} (X_i - \widehat{X}_i)^2 \qquad (4)$$

$$KL_{LV} = \frac{1}{n} \sum_{i}^{n} \frac{1}{2} \left[-\sum_i (\ln \sigma_i^2 + 1) + \sum_i \sigma_i^2 + \sum_i \mu_i^2 \right] \qquad (5)$$

In this study, the loss function is improved through an understanding of the physical leakage detection problem that the proposed framework attempts to solve. The analyses discussed in Section 4 establish that the leakages are associated with periods of anomalous flow (detected as outliers). Thus, to properly detect any leakages/bursts, it is necessary that the LKG and NLKG groupings possess different characteristics. While the differences in characteristics can be challenging to identify in the original time series domain (as discussed in Section 1), exaggerated differences in LV space (which is a sufficient and efficient representation of the original flow) can significantly improve the detection process. It is, therefore important that the VAE is able to accurately capture the distinction between LKG and NLKG groupings in the LV space.

As a remedial measure, this study uses an additional "domain-informed" loss term that drives the separation between the LVs of the two classes (i.e., LKG and NLKG). This is done mainly by computing the KL divergence (KL_{sep}) between the multivariate normal distributions of the LVs corresponding to the two classes (i.e., LKG and NLKG) as given in Eq. (6) where m_1 and m_2 and Σ_1 and Σ_2 are mean vectors and covariance matrices corresponding to the two classes of LVs, and n represents the number of groups (Hershey and Olsen, 2007). As can be understood from KL_{sep}, larger values of this term signify higher separation between the multivariate LV distributions of the two classes while lower values represent higher degree of overlap. Hence unlike KL_{LV} where the goal is minimizing the difference between the LV space and target distribution (hence lower values are better), the objective of the KL_{sep} loss is having higher values representing better separation and distinction between the two classes of LVs (corresponding to LKG and NLKG). Therefore KL_{sep} is added to the total loss of the VAE in an inverse manner as shown in Eq. (7).

$$KL_{sep} = \frac{1}{2} \left[(m_2 - m_1)^T \Sigma_2^{-1} + \text{tr}\left(\Sigma_2^{-1} \Sigma_1 \right) - \ln \frac{\det|\Sigma_1|}{\det|\Sigma_2|} - n \right] \qquad (6)$$

$$Loss_{Total} = Loss_{recon} + KL_{LV} + \frac{1}{KL_{sep}} \qquad (7)$$

Thus, the overall loss function used to train the VAE penalises three

items: i) improper reconstruction of the input sequence, ii) unregularised LV space, and iii) inseparable LVs across the two classes. This helps the VAE training process create distinct groupings in the LV space for the two classes, thereby improving confidence in LKG/NLKG classification. Alternative loss function additions, including computing the distance between class centroids, calculating class overlap probability, and finding the margins of SVM classifiers, were also explored during the internal training trials of domain-informed VAE. Based on the performance and consistency of implementation, finally KL_{sep} loss was selected. KL_{sep} is also compatible with KL_{LV} loss that is inherent to the VAEs. Hence, the purpose of the VAE in this study is to produce an LV mapping that shows the separation between the LVs of different types of flow time series groupings (LKG and NLKG), by capturing the different characteristics of these data groupings via dimensionality reduction.

For training the VAE in this study, the LKG and NLKG groupings are standardised and randomly split into train and test sets such that 80 % of both groupings are used in training (with 10 % cross-validation), and 20 % of both groups are used for testing the VAE. Various configurations of the VAEs were trained through grid search and hyperparameter tuning approaches (Briers et al., 2009, Mounce et al., 2013), to select the best-performing VAE architecture. The hyperparameter variations consisted of different: numbers of layers, number of neurons, activation functions, optimization algorithms, batch sizes, epochs, and dropout rates. The final optimised VAE is presented in Fig. 8. The proposed VAE consists of nine layers in each encoder and decoder (including the input and output layers) with a total of 1244 neurons and a bottleneck to produce two independent, normally distributed LVs. The activation function for each layer is hyperbolic tangent (tanh) except for the output layer of the decoder, which is linear (Sharma et al., 2020). The train set is shuffled into mini batches of 128 and used to train the VAE in 500 epochs using the adaptive moment estimation (Adam) (Kingma and Ba., 2015) optimiser and early stopping (Chollet, 2018) regularisation.

5.2. Support vector machine (SVM)

SVM is a supervised machine learning algorithm that can be applied to both regression and classification problems (Cortes and Vapnik, 1995). Since this study aims to train a model capable of accurately detecting the separation between the two classes of LVs, a binary SVM is deemed sufficient. The binary SVM is a linear classifier that, given training data and corresponding class labels, finds an optimal boundary in the feature space to maximise the separation between two classes. This boundary is called the optimal hyperplane (Boser et al., 1992). SVM classifiers identify the points closest to the hyperplane as support vectors. The support vectors influence the position and orientation of the optimal hyperplane. By maximising the thickness of the hyperplane (thereby distance between the support vectors), SVM allows the feature space to be divided into regions that represent the known classes. A hyperplane can be described by Eq. (8). The optimal hyperplane given in Eq. (8) is obtained through the optimisation of Eq. (9). Real data often contains outliers, and, thus is rarely linearly separable, so a soft margin SVM adds slack variables and regularisation to deal with noisy data (Xiao et al., 2019). Once the SVM has been trained and the hyperplane is
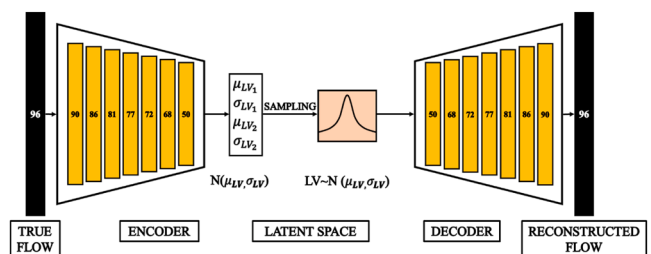

Fig. 8. The selected architecture of VAE (number of neurons of each layer is displayed in the cells).

obtained, new unlabelled data can be probabilistically classified by mapping into the feature space and noting position relative to the hyperplane.

$$w \cdot V + b = 0 \tag{8}$$

where w is the weight vector, b is the bias, and V is the input data.

$$min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i}^{n} \zeta_i \tag{9}$$

Subject to $k_i(w \cdot V_i + b) \geq 1 - \zeta_i$, with $\zeta_i \geq 0$, for all $1 \leq i \leq n$. Where n is the total number of input samples, $\| . \|$ is the matrix norm, and $C > 0$ is the regularisation constant. ζ is the slack variable, with $\zeta_i = 0$ for regular points and $\zeta_i > 0$ for outlier points. k is a variable such that negative classes have $k = -1$ and positive classes have $k = 1$.

The training of the SVM hyperplane is conducted using the LVs of the training dataset along with their associated LKG/NLKG class labels. From the ~8000 samples (80% of the ~10,000 groupings) in the training dataset, 383 points are selected as support vectors by the SVM, which are used to maximise the margin of the classifier and thereby classify the LV data. The SVM is used after the trained VAE encoder maps the flow grouping data into the LV space in order to utilise the trained hyperplane to probabilistically separate the LVs of the LKG and NLKG classes. Hence, once the proposed framework is provided with unlabelled 96×1 water flow data, it maps the flow onto the two-dimensional LV space using the pre-trained VAE encoder. And then, the pre-trained SVM uses the mapped LVs to determine the probability of the input flow data being classified as LKG and NLKG.

6. Results and discussion

This section presents the results of the trained framework on the train and test datasets of the flow time series. Both datasets contain one-third LKG groupings and two-thirds NLKG groupings so as to avoid any biases between the two groups. It should be noted that the framework is only trained using the train set and the test set acts as unseen data to assess the framework in this section. The following sections discuss the efficacy of LVs obtained from the VAE and the accuracy of the trained SVM.

6.1. VAE

The dataset utilised in this study contains leakages detected through outliers, which can differ significantly in magnitude. Since such different outliers are used to train the VAE (hence the LVs), it is important to assess the nature of the LVs with respect to the magnitude of outliers. However, due to the different diameters of pipes and flow rates for different DMAs, an absolute magnitude of the flow rate cannot be directly used to understand the corresponding LVs. Hence in this study, a standardised measure of volatility (i.e., the difference in the magnitude of the LKG flow compared to the magnitude and variance of preceding NLKG flow) is computed for each LKG grouping. This is done by computing the Z values for each of the LKG and NLKG groupings using Eq. (10). In this equation, $peak_{cur}$ represents the largest flow data point of the LKG/NLKG section and med_{pre} and σ_{pre} are the median value and standard deviation of the preceding NLKG flow data, respectively. The Z value thus compares the magnitude of the peak of the flow sequence to the average magnitude and variability of the preceding flow data, thereby indicating the magnitude of the outlier. The comparison of the Z values against the corresponding LVs for each grouping provides insights into the LVs. The means of the LVs (μ_{LV_1} and μ_{LV_2}) trained using the VAE are presented in Fig. 9, with each point colour coded as either LKG or NLKG along a colour gradient set by the corresponding Z values.

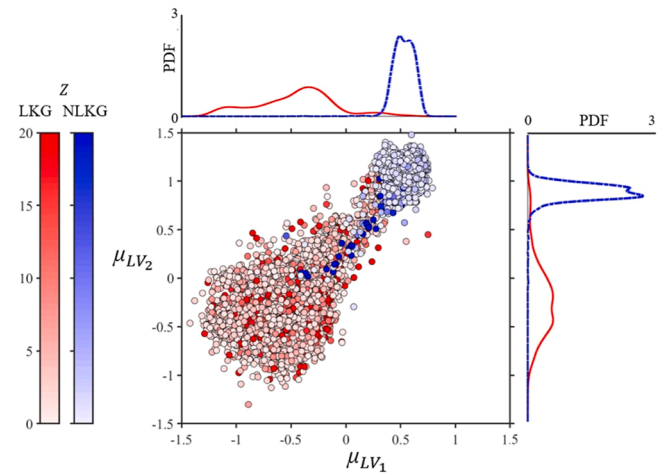$$Z = \frac{peak_{cur} - med_{pre}}{\sigma_{pre}} \tag{10}$$

Fig. 9. Mean LVs (μ_{LV_1} and μ_{LV_2}) of the trained VAE.

Fig. 9 shows clear groupings of μ_{LV_1} and μ_{LV_2} for LKG and NLKG data. There is minimal overlap in the probability density functions (PDF) distributions of the two groupings for both mean LVs, with a distinct and dense cluster of NLKG LVs and a largely separate, though more spread, the cluster of LKG LVs. The difference in the spread of the clusters can be accredited to the fact that the LKG groupings are generally uniform (Fig. 7b) and obviously fluctuate less dramatically than the outliers NLKG groupings (Fig. 7a). Hence the corresponding mean LVs are narrowly spread for the NLKG groupings. A significant majority of NLKG data leads to μ_{LV_1} between 0 and 1, and μ_{LV_2} between 0.5 and 1.5. The PDF kernels for the mean LVs of NLKG show sharp spikes in these ranges, indicating a dense concentration and smaller variance. As mentioned earlier, this narrow variance in the LV space suggests minimal variation in the characteristics of the initial time series data of the NLKG groupings. The dimensionality reduction leads to similar features in the input data. The Z values are represented in the red and blue colour gradients for the LKG groupings and NLKG groupings respectively. The mean LVs for the LKG groupings show no major difference between those with high Z values, indicating higher volatility in the outliers, and those with low Z values. The NLKG groupings lead to low Z values, as expected for non-outlier data, and are shown in the dense groups of pale blue points. The few exceptions are observed for the NLKG groupings with particularly high Z values, indicating greater volatility in the original time series than the majority of NLKG groupings. An examination of these NLKG groupings characterised by elevated Z values reveals that the corresponding flow time series data typically exhibits significantly diminished magnitudes. This phenomenon can be attributed to several factors, including variations in DMA size, pipe diameter, and potential supply constraints. The pronounced sensitivity of Z values to even minor fluctuations in low-flow scenarios results in their escalation. Detecting leakage within pipes of lower volume can consequently pose a more intricate task compared to their higher-volume counterparts, although the magnitude of water loss associated with leaks in lower volume pipes is anticipated to be comparatively minimal.

The μ_{LV_1} and μ_{LV_2} corresponding to LKG data have a greater spread, with PDF kernels showing lower peaks and higher variances. This can be due to the higher variability in the characteristics of the input time series of the LKG groupings. The spread is observed to not differ significantly for the different Z values. This suggests that the volatility of LKG groupings (relative to preceding non-leakage flow) is not the only characteristic influencing the dimensionality reduction process. Hence, the training of VAEs leads to surrogate LVs that may capture other behaviours of the LKG time series (e.g., the sustained value of peaks, etc.). However, such analysis is out of the scope of this study. The possible reasons behind the few misclassified LKG groups are not immediately discernible. However, it is noteworthy that these instances largely fall

within the margins of the SVM classifier.

The reconstruction power of the VAE is assessed using the index of agreement (*IA*) (Willmott, 1981, Willmott et al., 1985) described in Eq. (11), where X is the true input data, $\widehat{X}$ is the reconstructed data, and n and i represent the total number of timesteps and the timestep of interest, respectively. *IA* is computed by comparing the input flow time series (X) and the output flow time series ($\widehat{X}$) reconstructed through the decoder of the VAE using the LVs. *IA* gives a single bounded metric for pattern characterisation (value of 1 represents perfect match and 0 represents no match between the true and reconstructed flow time series) and comparison, and also incorporates information of the magnitude of deviations. *IA* is, therefore, a valuable tool for the comparison of model performance and has been widely applied to the assessment of time series models (Durbin and Koopman, 2012). It is worth noting that the primary goal of the proposed framework is to train surrogate LVs that lead to an accurate classification of LKG and NLKG flow groupings through SVM, and therefore a precise reconstruction of the input groupings is not the main goal of the study and only serves to improve confidence in the surrogate LVs. Hence, *IA* of the ~10,000 flow groupings provide an additional indicator of the strength of model performance.

$$IA = 1 - \frac{\sum_{i=1}^{n}(X_i - \widehat{X}_i)^2}{\sum_{i=1}^{n}(|\widehat{X}_i - \text{mean}(X)| + |X_i - \text{mean}(X)|)^2} \qquad (11)$$

Fig. 10 shows the *IA* distributions for both the LKG and NLKG groupings. The *IA* values of both LKG and NLKG groupings follow similar distributions, with an *IA* > 0.5 for most cases for both train and test groupings. The median *IA* value for LKG groupings is 0.67, while the median value for NLKG groupings is 0.59. This indicates that the VAE is able to handle the higher variability of LKG groupings well. The train and test datasets follow almost identical IA distributions, with lower quartiles of 0.51 and 0.52, respectively, and both have a median of 0.61 and an upper quartile of 0.73. With less than 25 % of the dataset having an *IA* value below 0.5, the VAE is verified as having sufficient reconstruction power. The LVs are, therefore, successful in providing sufficient information for reconstructing time series data and showing the separation between LKG and NLKG classes. Given that, in this study, dimensionality is reduced from 96 points to a two-dimensional LV space, it is impressive that the VAE can produce LVs that sufficiently achieve both of these aims in only a two-dimensional LV space.

6.2. SVM

A radial basis kernel-based SVM binary classifier is trained on the mean LVs (μ_{LV_1} and μ_{LV_2}) obtained from the VAE using the training set. The SVM is trained to classify the mean LVs into LKG and NLKG accurately. Hence the LVs are the inputs to the SVM classifier, and the associated LKG/NLKG labels are the targets. The hyperplane and associated margins (ε) of the trained SVM are shown in the LV space in
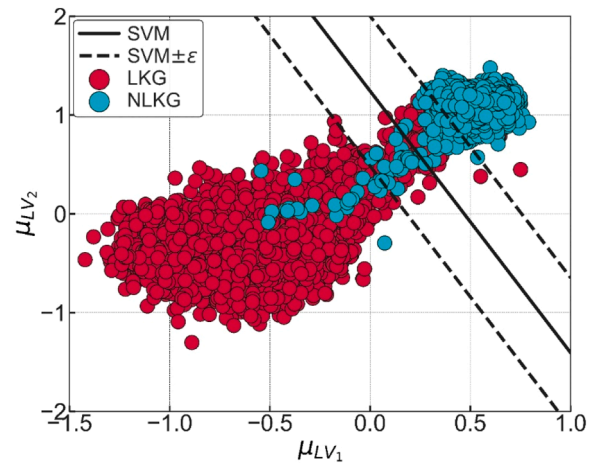

Fig. 11. SVM hyperplane and margins (ε) shown in latent variable space.

Fig. 11. It can be observed that the hyperplane seeks to create the largest possible separation between the two classes of LKG and NLKG, with the margins largely covering the area of overlap between the two classes. Beyond these margins, only a few points are incorrectly classified.

The performance of the trained SVM is then tested on the test dataset. This dataset consists of μ_{LV_1} and μ_{LV_2} and their associated LKG/NLKG labels corresponding to the flow time series in the test dataset (2031 examples with two-thirds of NLKG and one-third of LKG data). The confusion matrix and receiver operating characteristic (ROC) curve of the classification results are presented in Fig. 12a, b, respectively. It can be observed from Fig. 12a that the SVM leads to an overall accuracy of 98.2 % on the test set for classifying the LVs into LKG and NLKG classes. Furthermore, the SVM's precision represents the fraction of LKG predictions in the LKG class, and the recall of the SVM, which is the fraction of all LKG LV inputs that were correctly predicted as LKG, is 98.3 % and 95.9 %, respectively. The trained SVM further leads to an F1 score (which combines precision and recall into a single metric by calculating their harmonic mean) of 97.1 %. .

The ROC represents a probability curve that provides a measure of how well a model can separate two classes for different thresholds. The area under the ROC curve (*AUC*) indicates classification accuracy and can range from a minimum of zero to a maximum of one. Fig. 12b shows the ROC-AUC curve for the trained SVM classifier. It can be observed that the SVM classifier leads to a high AUC value of 0.996, thereby indicating its excellent classification power. Furthermore, the strong performance on the test dataset demonstrates that the SVM could be used to accurately classify any new, unlabelled time series groupings as either LKG or NLKG, based on the mapping of the corresponding LVs onto LV space.
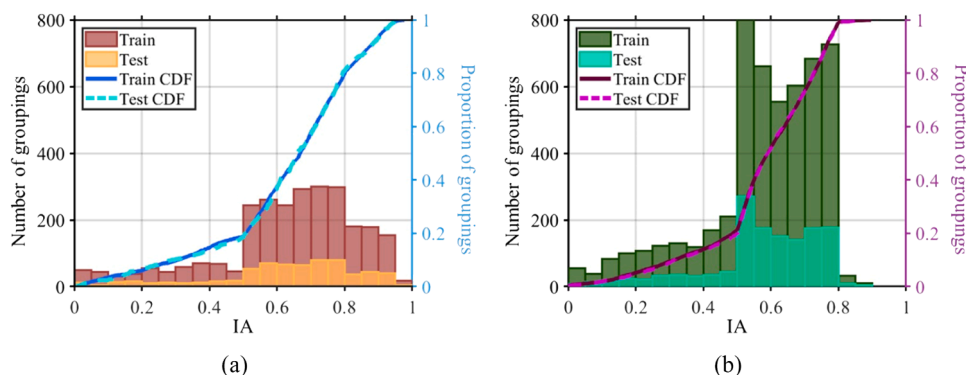

(a)	(b)

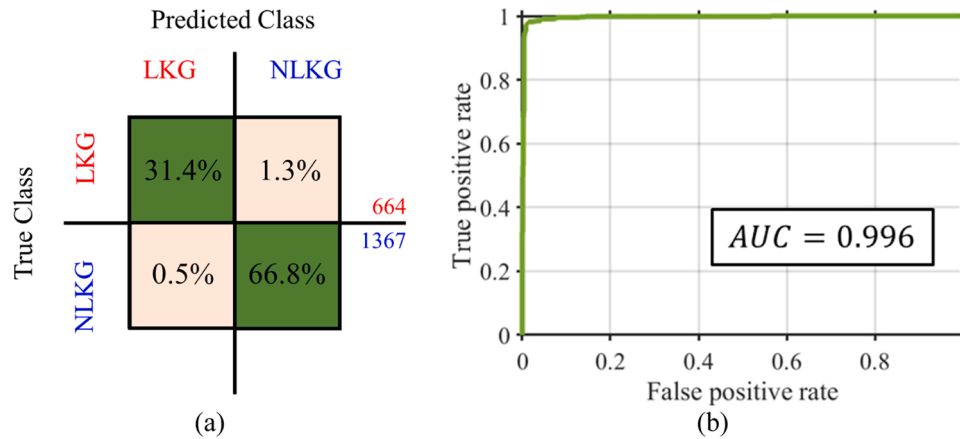**Fig. 10.** The *IA* distributions of (a) LKG and (b) NLKG groupings for both train and test sets.

Fig. 12. The SVM (a) confusion matrix (for the 2031 points in the test dataset) and (b) ROC curve for the SVM classifier, with $AUC = 0.996$.

7. Comparison with MNF metric

To assess the performance of the proposed framework against a traditional leakage detection method, a simplified MNF analysis is conducted for the dataset. While the available data is not highly sufficient to carry out a comprehensive MNF analysis, there is precedence for using a simplified MNF metric based solely on flow data (Mounce et al., 2007). A simplified MNF index finds the ratio of MNF for a given night to the mean or median of MNF over a preceding window period. High MNF index values are more likely to represent abnormal flow events, which include leakage. The size of the window period varies in the literature from three days to six months (Amoatey et al., 2021; Lee et al., 2022; Tabesh et al., 2009; Huang et al., 2018), though it is generally agreed that a larger window can give a better representation of typical night flow behaviour for a given DMA. This study uses a window of seven days, as this covers a full week of flow data (including the weekend) yet is short enough that sufficient data is available for most of the groupings. The nighttime hours used to calculate MNF also vary in literature, beginning as early as 12am and ending as late as 5am (Lee et al., 2022). This study uses the hours of 2am to 4am, which are selected in various existing studies (Yu and Zhu, 2020). To find the MNF index, the median value of flow during these hours (night flow, NF) over the window period is found. The median is chosen over the mean to limit the effects of any erroneous data, as MNF is sensitive to fluctuations or anomalies. For the night of interest, a significant deviation from this value during the same hours can be taken to indicate possible leakage under MNF analysis. This deviation is computed using Eq. (12), where d is the 24-hour day of interest and i is the size of the window period in days.

$$\text{MNF index} = \frac{\text{median}(\text{NF}_d)}{\text{median}(\text{NF}_{d-1},\ \text{NF}_{d-2},\ \dots,\ \text{NF}_{d-i})} \qquad (12)$$

For the purpose of comparison, this study defines an MNF index of 1.1 or greater, which represents >10 % deviation from the median preceding NF, as indicative of potential leakage. This aligns with MNF index values observed in the literature (Mounce et al., 2007), however this threshold could be adjusted if necessary. Hence based on this, the MNF index is computed for the ~10,000 train and test flow groupings.

To assess the accuracy of this MNF index for leakage identification on the dataset used for this study, the train and test datasets are combined so that all ~10,000 LKG and NLKG groupings are analysed. The confusion matrix of the classification results of the MNF index analysis are presented in Fig. 13. Of the 10,000+ groupings, the MNF analysis accurately classifies 70.7 %. Though the MNF tends to perform satisfactorily for a complex problem like leakage detection, it fails to match the accuracy of the proposed VAE-SVM framework on this dataset. An additional benefit of the framework over the MNF index is that the framework does not require a specific period of overnight flow or other
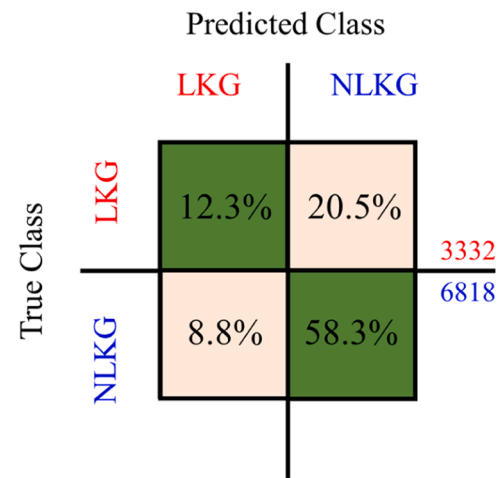

Fig. 13. Confusion matrix based on MNF Index.

similar assumptions. This allows the framework to be flexible and identify possible leakages at any time of day, allowing a more rapid identification.

8. Conclusion

In developed urban water networks, reducing leakage levels is not only crucial to address the significant costs incurred by the sector but also to promote sustainability and build resilience in the water supply chain. Leakage levels of over 20 % not only pose a financial burden (PR19 final determinations 2019) but also affect consumer confidence in the reliability of their water supply, hindering the widespread adoption of sustainable water conservation practices. While large-scale upgrades to the UK's aging water distribution infrastructure will take time, there are opportunities now to make huge improvements in how this infrastructure is monitored and managed, leading to faster identification and repair of leaks (McMillan and Varga, 2022, '"Proactive" approach to leaks required to meet tough Ofwat targets 2023). By leveraging machine learning algorithms to analyse time series data provided by flow sensors, accurate detection of leaks at the DMA level can be achieved without costly sensor upgrades. This strategic approach to upgrading and repairing infrastructure not only offers significant economic benefits but also promotes sustainability by better utilising existing sensing capabilities to reduce water and energy usage.

This study proposes a hybrid machine-learning based framework for rapid classification of incoming water flow time series data. The

framework consists of a domain-informed VAE, which is trained using a loss function that mathematically recognises that the characteristics of leakage flow should be different from those of regular (non-leakage) flow. After the VAE encoder reduces the dimensionality of the time series data into two surrogate LVs, a binary SVM classifier is then used to create a hyperplane to separate the LVs of the two classes. Once trained on examples of both regular (non-leakage) and leakage flow, the proposed framework is able to classify unlabelled flow data as leakage or non-leakage with a high degree of accuracy. The data-driven framework is trained and tested on a dataset of 12-months of flow and repair data for over 2500 DMAs managed by Yorkshire Water, UK. The data is carefully pre-processed and appropiately sampled to obtain ~10,000 flow time series groupings (out of which 66.66% groupings are NLKG and 33.33% groupings are LKG) of 96x1 dimension (representing 24 hours of flow). The framework is trained on randomly sampled ~8000 training examples of LKG and NLKG groupings and then tested on the remaining dataset of over 2000 unseen flow groupings. The framework is able to classify the test dataset groupings with an accuracy of 98.2 %. Furthermore, an AUC value of 0.996 is observed, highlighting the strong classification power of the proposed VAE-SVM framework.

Though this study uses water flow time series data, it is worth recognising that the pressure data is also collected in many water distribution networks. Future research endeavours could potentially benefit from integrating both flow and pressure time series as inputs to the models. This integration might enhance the accuracy of leakage detection by providing a more comprehensive view of the network's dynamics. Additionally, the current study operates under the assumption that the historical water flow data can effectively train a model for real-time leakage classification. This presupposes that historical non-leakage water consumption patterns remain indicative of current usage. However, given the significant shifts in consumption patterns, particularly following the COVID-19 pandemic, it becomes increasingly relevant to incorporate more recent data for non-leakage samples where available.

On the scientific front, it is important to acknowledge inherent limitations associated with neural networks and data-driven methods employed in this study. While powerful in pattern recognition, the extrapolative capabilities of neural networks have been a matter of debate in many cases (Pastore and Carnini, 2021). Additionally, data-driven approaches heavily rely on the quality and representativeness of the data used for training. In the context of water distribution networks, this means that if the training data does not adequately capture certain types of leakages or network conditions, the model's performance in real-world applications might be compromised. Another consideration is the interpretability of the trained models, which can be challenging due to their complex and often opaque nature (Somani et al., 2023, Fayaz, 2023). This can make it difficult to understand the underlying reasons for the model's predictions and decision-making, which is crucial for gaining user trust and acceptance in practical applications.

Furthermore, as water distribution networks increasingly adopt smart meters, new opportunities arise in leakage detection. These smart meters can serve as additional point sensors, potentially enhancing the granularity and accuracy of flow monitoring within DMAs. Consequently, there is scope for future research to extend the current method beyond the DMA level, aiming for more precise leakage localisation. Such advancements could leverage the additional data provided by smart meters to improve the model's accuracy and reliability in detecting and pinpointing leakages within the network.

Using Bayesian deep learning methods to analyse time series data from water flow sensors is a novel approach to leakage detection and this study demonstrates their high levels of classification accuracy. The proposed framework has the potential to improve the health monitoring of water distribution networks without requiring costly and time-consuming upgrades to the sensor network. The proposed framework is compatible with the current set-up of water distribution systems in the UK, and its automated nature can facilitate timely andf cost-efficient identification of potential leakage. By promptly detecting and

repairing leaks, the water supply can be better protected from disruptions, enhancing its reliability and resilience. The proposed approach can also reduce the potential for major water losses, which can strain the capacity of the distribution system and lead to costly repairs and downtime. Minimising these disruptions, as well as driving down the proportion of bursts identified by consumers, will build confidence in a continuous and efficient water supply, which will in turn promote sustainable demand-side behaviours. Hence, such efficient leakage identification represents an important step towards smart and sustainable urban water management (McMillan and Varga, 2022).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data source is open access and is given in manuscript. Code can be made available on request.

References

Abdelmageed, S., Tariq, S., Boadu, V., Zayed, T., 2022. Criteria-based critical review of artificial intelligence applications in water-leak management. Environ. Rev. 30 (2), 280–297. https://doi.org/10.1139/er-2021-0046.

Ahmad, Z., Nguyen, T.K., Kim, J.-M., 2023. Leak detection and size identification in fluid pipelines using a novel vulnerability index and 1-D convolutional neural network. Eng. Appl. Comput. Fluid Mech. 17 (1), 2165159 https://doi.org/10.1080/19942060.2023.2165159.

Aksela, K., Aksela, M., Vahala, R., 2009. Leakage detection in a real distribution network using a SOM. Urban Water J. 6 (4), 279–289. https://doi.org/10.1080/15730620802673079.

Amoatey, P.K., Obiri-Yeboah, A.A., Akosah-Kusi, M., 2021. Impact of active night population and leakage exponent on leakage estimation in developing countries. Water Pract. Technol. 17 (1), 14–25. https://doi.org/10.2166/wpt.2021.124.

Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A., Oliveira, M.S., 2018. Short-term water demand forecasting using machine learning techniques. J. Hydroinf. 20 (6) https://doi.org/10.2166/hydro.2018.163. Art. no. 6.

Asperti, A., Trentin, M., 2020. Balancing reconstruction error and kullback-leibler divergence in variational autoencoders. IEEE Access 8, 199440–199448. https://doi.org/10.1109/ACCESS.2020.3034828.

Bellman, R., 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press.

Birek, L., Petrovic, D., Boylan, J., 2014. Water leakage forecasting: the application of a modified fuzzy evolving algorithm. Appl. Soft Comput. 14, 305–315. https://doi.org/10.1016/j.asoc.2013.05.021.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. New York, NY, USA. Association for Computing Machinery, pp. 144–152. https://doi.org/10.1145/130385.130401. COLT '92.

Briers, M., Doucet, A., Maskell, S., 2009. Smoothing algorithms for state–space models. Ann. Inst. Stat. Math. 62 (1), 61. https://doi.org/10.1007/s10463-009-0236-2.

Chan, T.K., Chin, C.S., Zhong, X., 2018. Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. IEEE Access 6, 78846–78867. https://doi.org/10.1109/ACCESS.2018.2885444.

Chollet, F., 2018. Deep Learning with Python, 1st edition. Manning Publications, Shelter Island, New York.

Cody, R.A., Tolson, B.A., Orchard, J., 2020. Detecting leaks in water distribution pipes using a deep autoencoder and hydroacoustic spectrograms. J. Comput. Civil Eng. 34 (2), 04020001 https://doi.org/10.1061/(ASCE)CP.1943-5487.0000881.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3) https://doi.org/10.1007/BF00994018. Art. no. 3.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis By State Space Methods. OUP. Oxford.

Fayaz, J., Galasso, C., 2023. A deep neural network framework for real-time on-site estimation of acceleration response spectra of seismic ground motions. Comput. Aided Civ. Infrastruct. Eng. 38 (1), 87–103. https://doi.org/10.1111/mice.12830.

Fayaz, J., 2023. ROSERS—A deep learning framework for earthquake early warning and its interpretation'. In: Shrikhande, M., Agarwal, P., Kumar, P.C.A. (Eds.), in *Proceedings of 17th Symposium on Earthquake Engineering (Vol. 4)*, Lecture Notes in Civil Engineering. Springer Nature, Singapore, pp. 473–486. https://doi.org/10.1007/978-981-99-1459-3_38.

Feng, P., Shen, Y., Liu, W., 2021. Water leakage detection based on variation bayesian neural network autoencoder. J. Phys.: Conf. Ser. 1948 (1), 012110 https://doi.org/10.1088/1742-6596/1948/1/012110.

Fu, G., Jin, Y., Sun, S., Yuan, Z., Butler, D., 2022. The role of deep learning in urban water management: a critical review. Water Res. 223, 118973 https://doi.org/10.1016/j.watres.2022.118973.

García, V.J., Cabrera, E., Cabrera, J., 2008. The Minimum Night Flow Method Revisited. In: Proceedings of the Eighth Annual Water Distribution Systems Analysis (WDSA). Presented at the Eighth Annual Water Distribution Systems Analysis Symposium. American Society of Civil Engineers, Cincinnati, Ohio, USA, pp. 1–18. https://doi.org/10.1061/40941(247)35.

Geberemariam, T., Juran, I., Shahrour, I., 2014. Virtual DMA municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class SVM. J. Pattern Recognit. Res. 9, 25–42. https://doi.org/10.13176/11.548.

González-Muñiz, A., Díaz, I., Cuadrado, A.A., García-Pérez, D., Pérez, D., 2022. Two-step residual-error based approach for anomaly detection in engineering systems using variational autoencoders. Comput. Electr. Eng. 101, 108065 https://doi.org/10.1016/j.compeleceng.2022.108065.

Hershey, J.R., Olsen, P.A., 2007. Approximating the kullback leibler divergence between gaussian mixture models. In: Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07. https://doi.org/10.1109/ICASSP.2007.366913. IV-317-IV–320.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507. https://doi.org/10.1126/science.1127647.

Huang, P., et al., 2018. Real-time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning. Water 10 (12). https://doi.org/10.3390/w10121765 (Basel)Art. no. 12.

i2O Water - Pressure logging or acoustic logging?, 2020. https://en.i2owater.com/pressure-logging-or-acoustic-logging/ (accessed 3.13.23).

Jung, D., Lansey, K., 2015. Water distribution system burst detection using a nonlinear kalman filter. J. Water Resour. Plann. Manag. 141 (5), 04014070 https://doi.org/10.1061/(ASCE)WR.1943-5452.0000464.

Kang, J., Park, Y.J., Lee, J., Wang, S.H., Eom, D.-S., 2018. Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. IEEE Trans. Ind. Electron. 65 (5), 4279–4289. https://doi.org/10.1109/TIE.2017.2764861.

Kingma, Ba, J., 2015. Adam: A Method for Stochastic Optimization, 3rd. San Diego, California, USA. https://doi.org/10.48550/arXiv.1412.6980.

Kingma, D.P., Welling, M., 2019. An introduction to variational autoencoders. Found. Trends Mach. Learn. 12 (4), 307–392. https://doi.org/10.1561/2200000056.

Kullback, S., Leibler, R.A., 1951. On Information and Sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Lee, S.S., Lee, H.H., Lee, Y.J., 2022. Prediction of minimum night flow for enhancing leakage detection capabilities in water distribution networks. Applied Sciences 12 (13). https://doi.org/10.3390/app12136467. Art. no. 13.

Leu, S.-S., Bui, Q.-N., 2016. Leak prediction model for water distribution networks created using a bayesian network learning approach. Water Resour. Manag. 30 (8), 2719–2733. https://doi.org/10.1007/s11269-016-1316-8.

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2008. Isolation forest. In: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17.

Liu, F.T., Ting, K.M., Zhou, Z.H., 2012. Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data 6 (1). https://doi.org/10.1145/2133360.2133363, 3:1-3:39.

Marlow, D.R., Moglia, M., Cook, S., Beale, D.J., 2013. Towards sustainable urban water management: a critical review. Water Res. 47 (20), 7150–7161. https://doi.org/10.1016/j.watres.2013.07.046.

Masreliez, C., Martin, R., 1977. Robust bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Automat. Contr. 22 (3), 361–371. https://doi.org/10.1109/TAC.1977.1101538.

McMillan, L., Varga, L., 2022. Towards self-healing in water infrastructure systems. In: Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction, pp. 1–9. https://doi.org/10.1680/jsmic.22.00006.

Morrison, J., 2004. Managing leakage by district metered areas: a practical approach. Water 21, 44–46 (Basel).

Mounce, S.R., Machell, J., 2006. Burst detection using hydraulic data from water distribution systems with artificial neural networks. Urban Water J.l 3 (1), 21–31. https://doi.org/10.1080/15730620600578538.

Mounce, S., Boxall, J., Machell, J., 2007. An artificial neural network/fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation. In: Proceedings of the Water Management Challenges in Global Change, p. 320.

Mounce, S.R., Boxall, J.B., Machell, J., 2010. Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. J. Water Resour. Plann. Manag. 136 (3) https://doi.org/10.1061/(ASCE)WR.1943-5452.0000030. Art. no. 3.

Mounce, S.R., Mounce, R.B., Jackson, T., Austin, J., Boxall, J.B., 2013. Pattern matching and associative artificial neural networks for water distribution system time series data analysis. J. Hydroinf. 16 (3), 617–632. https://doi.org/10.2166/hydro.2013.057.

Pandey, P., Bokde, N.D., Dongre, S., Gupta, R., Feb. 2021. Hybrid models for water demand forecasting. J. Water Resour. Plann. Manag. 147 (2), 04020106 https://doi.org/10.1061/(ASCE)WR.1943-5452.0001331.

Pastore, A., Carnini, M., 2021. Extrapolating from neural network models: a cautionary tale. J. Phys. G: Nucl. Part. Phys. 48 (8), 084001 https://doi.org/10.1088/1361-6471/abf08a.

'"Proactive" approach to leaks required to meet tough Ofwat targets, 2023. Glob. Water Intell. 21 (7). Jul. 23, 2020. Accessed: Feb. 16[Online]. Available: https://www.globalwaterintel.com/global-water-intelligence-magazine/21/7/smart-water-watch/proactive-approach-to-leaks-required-to-meet-tough-ofwat-targets.

PR19 final determinations, 2019. Securing cost efficiency technical appendix. Ofwat. Accessed: Jun. 29, 2021. [Online]. Available: https://www.ofwat.gov.uk/wp-content/uploads/2019/12/PR19-final-determinations-Securing-cost-efficiency-technical-appendix.pdf.

Pu, Z., et al., 2023. A hybrid Wavelet-CNN-LSTM deep learning model for short-term urban water demand forecasting. Front. Environ. Sci. Eng. 17 (2) https://doi.org/10.1007/s11783-023-1622-3.

Puust, R., Kapelan, Z., Savic, D.A., Koppel, T., 2010. A review of methods for leakage management in pipe networks. Urban Water J. 7 (1), 25–45. https://doi.org/10.1080/15730621003610878.

Romano, M., Kapelan, Z., Savić, D.A., 2014. Automated detection of pipe bursts and other events in water distribution systems. J. Water Resour. Plann. Manag. 140 (4), 457–467. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000339.

Sharma, S., Sharma, S., Athaiya, A., 2020. Activation functions in neural networks. IJEAST 04 (12), 310–316. https://doi.org/10.33564/IJEAST.2020.v04i12.054.

Shu, X., Bao, T., Zhou, Y., Xu, R., Li, Y., Zhang, K., 2023. Unsupervised dam anomaly detection with spatial–temporal variational autoencoder. Struct. Health Monit. 22 (1), 39–55. https://doi.org/10.1177/14759217211073301.

Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the Em algorithm. J. Time Ser. Anal. 3 (4), 253–264. https://doi.org/10.1111/j.1467-9892.1982.tb00349.x.

Somani, A., Horsch, A., Prasad, D.K., 2023. Interpretability in Deep Learning. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-20639-9.

Tabesh, M., Yekta, A.H.A., Burrows, R., 2009. An integrated model to evaluate losses in water distribution systems. Water Resour. Manag. 23 (3), 477–492. https://doi.org/10.1007/s11269-008-9284-2.

Taylor, S., Hodges, J., 2008. Water supplies for Auckland - predict and provide or conserve and contain? Water Pract. Technol. 3 (1), wpt2008023 https://doi.org/10.2166/wpt.2008.023.

W. Todo, B. Laurent, J.-M. Loubes, and M. Selmani, 'Dimension reduction for time series with variational autoencoders, 2022. doi:10.48550/arXiv.2204.11060.

United Utilities - World's biggest listening project helps tackle water leaks [WWW Document], 2019. United Utilities. https://www.unitedutilities.com/corporate/newsroom/latest-news/worlds-biggest-listening-project-will-help-save-water/ (accessed 3.22.23).

van der Maaten, L., Postma, E., van den Herik, J., 2009. Dimensionality reduction: a comparative review. J. Mach. Learn Res. 10, 66–71.

Verleysen, M., François, D., Cabestany, J., Prieto, A., Sandoval, F., 2005. The curse of dimensionality in data mining and time series prediction', in computational intelligence and bioinspired systems. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 758–770. https://doi.org/10.1007/11494669_93.

Wang, C., Han, F., Zhang, Y., Lu, J., 2020. An SAE-based resampling SVM ensemble learning paradigm for pipeline leakage detection. Neurocomputing 403, 237–246. https://doi.org/10.1016/j.neucom.2020.04.105.

Willmott, C.J., et al., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res.: Oceans 90 (C5), 8995–9005. https://doi.org/10.1029/JC090iC05p08995.

Willmott, C.J., 1981. On the Validation of Models. Phys. Geogr. 2 (2), 184–194. https://doi.org/10.1080/02723646.1981.10642213.

Xiao, R., Hu, Q., Li, J., 2019. Leak detection of gas pipelines using acoustic signals based on wavelet transform and Support Vector Machine. Measurement 146, 479–489. https://doi.org/10.1016/j.measurement.2019.06.050.

Ye, G., Fenner, R.A., 2011. Kalman filtering of hydraulic measurements for burst detection in water distribution systems. J. Pipeline Syst. Eng. Pract. 2 (1), 14–22. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000070.

Yeo, D., Bae, J.H., Lee, J.C., 2019. Unsupervised learning-based pipe leak detection using deep auto-encoder. J. Korea Soc. Comput. Inf. 24 (9), 21–27. https://doi.org/10.9708/jksci.2019.24.09.021.

Yu, T., Zhu, H., 2020. Hyper-Parameter Optimization: A Review of Algorithms and Applications. https://doi.org/10.48550/arXiv.2003.05689.

Zavrak, S., İskefiyeli, M., 2020. Anomaly-based intrusion detection from network flow features using variational autoencoder. IEEE Access 8, 108346–108358. https://doi.org/10.1109/ACCESS.2020.3001350.

Zhou, X., et al., 2019. Deep learning identifies accurate burst locations in water distribution networks. Water Res. 166, 115058 https://doi.org/10.1016/j.watres.2019.115058.