



Virtual DMA Municipal Water Supply Pipeline Leak Detection and Classification Using Advance Pattern Recognizer Multi-Class SVM

Thewodros G. Mamo

*Department of Civil & Urban Engineering
New York University Polytechnic School of Engineering
6 Metro Tech Center, Brooklyn, NY 11201*

tgm232@nyu.edu

Ilan Juran

*Department of Civil & Urban Engineering
New York University Polytechnic School of Engineering
6 Metro Tech Center, Brooklyn, NY 11201*

ijuran@poly.edu

Isam Shahrouf

*Laboratoire de mécanique de Lille,
University of Sciences and Technologies of Lille,
59650 Villeneuve d'Arcy, France*

isam.shahrouf@univ-lille1.fr

Abstract

In this paper we investigated and analyzed the concept of virtual district metered area (DMA) as the core objective of the research to resolve the current gap and limitations of the actual district metered area state of practice through the development of virtual district metered area pipeline leak detection and classification system using multi-class support vector machine (SVM) advanced pattern recognizer at Lille University water supply pipeline networks study area the so called "Zone-6". The SVMs were trained on multiple cases representing the presence of leaks in various sizes and locations. The research results, and analysis showed a rather promising performance, which could be successfully implemented. Moreover, the proposed method could enable the water utility companies and other stakeholders to further reduce risks associated with pipeline leaks or breaks. This method also can be used during decision-making process for selecting which pipeline requires urgent action, and engineer the optimal short-term response or alternative for maintenance strategies. Furthermore, the proposed methodology could benefit the water utility companies by reducing the cost and operational drawbacks associated with implementing the actual district metered area (DMA). It also improve the day to day operational decision making process by detecting and classifying the different stages of pipelines leaks and breaks according to their severity, which can enable the operators to see the behavior of the network on the control room screens they are familiar with and enable them to quickly perform the best short term response strategy.

Keywords: Water Utility, Infrastructure, EPANET, Multi-class SVM, Pattern Recognition, Leaks Detection, DMA, Virtual DMA, Posterior Probability.

1. Introduction

Municipal water distributions pipelines leaks are serious problems for many big cities, and remain a tremendous challenge for public water utility companies. The economic and property damage costs associated with pipeline leaks and breaks are rapidly rising in an alarming rate and become the main causes for loss of revenue. Undetected leaks could continually increase the amount of water loss without being invoiced or metered. Water losses deprive public water utility companies from additional revenue and paralyzed their effort to expand

services. The overall impairment caused by leaks and breaks depend on the time between the actual leak detected and its isolation. For many of big city water utility companies these tasks are complicated due to lack of active leaks detection & monitoring system. So the presences of leaks are identified when it became visible on the surface, by this time the consequences can be both expensive and harmful. Therefore, these challenges keep the water utility companies actively searching for innovative approaches to early detection and classification methods of leaks and [1–3, 11].

2. DMA Methodology for Leak Detection and the Challenges

In the early 1980s the UK water utility companies has introduced the concept of managing WS pipeline network using DMA [13, 29, 30]. The DMA methodology for pipeline leaks monitoring and detection requires careful design and installation of flow meters and closing valves at strategic points throughout the WS pipeline networks. This process, which creates a well-defined sub area or zone out of a big water distribution pipe line network, is called a district metered area (DMA), and used to control pressure and monitor leaks in water supply networks.

The practical cost effective and efficient leaks management analysis using DMA depends on the type and sophistication level data analysis capabilities and associated system in-place like SCADA by water utility companies to identify those areas in the network, with considerable anomalies which have potential and considered to be the major contributor of pipeline failure or volume of water loss [7, 29, 30]. The principal advantage of DMA is that the key characteristics of WDS such as pressure, demand, and water quality of well-defined area of the distribution system can be closely monitored [15, 16, 30]. DMA allows the water utility companys managers to prioritize the most cost effective maintenance strategies. This approach also provides a better knowledge of how the system works and enables to manage pressure, monitoring and investigation of leaks. In general, by creating DMAs, the water utility companies face the following the challenges [20–26].

- Cascading DMAs:- these challenges are common when DMAs are established having two or more metering location sites that water first pass through one or more other sub DMAs before entering the designated DMA. The challenge under these conditions is to seize the flow rate data within the same time frame which is important for accurate calculation and analysis of instantaneous flow rates.
- Water quality challenge: - to establish DMAs, the pipeline network systems valves have to be closed along DMAs boundaries, which result in increasing considerably the number of dead ends, which can create the possibility of taste and odor problems with low chlorine residuals, particularly in areas with branch pipeline system, which will create accumulation of debris resulting in discoloration or even blockages. Even if some of the solution such as routine flushing, valve operation, or fixed rate jumpers across boundary valves can be implemented, where water quality is an issue or consumer complaints arise it can overcome the cost aspect of challenge and put the entire proposed DMA cost effectiveness in question.
- Less robust under failure conditions challenges, most open network systems automatically compensate (up to a point) for changes in demand patterns. DMAs, on the other hand need to be managed to allow for mains improvement, peak demands, loss of supply etc.

- The costs of establishment can be considerable, not only are meters and data loggers required, but also new and replacement valves may be needed. In some cases, fresh tracing and mapping of the network system may be necessary.
- A substantial commitment also required from management and workforce, because It is vital that valves are checked and meters read regularly, otherwise the information obtained is misleading or useless. This too has a cost, which has to be accepted and budgeted. Therefore, the core objective of this research paper is to resolve the above current gap and limitations of the DMA state of practice through the development of virtual DMA.

3. Virtual DMA

Virtual DMA can be defined as the concept of monitoring and identifying leaks throughout the entire WS pipeline networks without creating actual DMA or Sub-DMAs, instead multi-parameter sensor installed at key positions used to create “virtual DMA zones”. in this new methodology water utility companies (WUC) can use multi-parameter monitoring technology composed of sensors that can simultaneously measure bidirectional flow, pressure flow rate, volume of consumption etc. record data and have capability of communicate with others device in place using statistical machine learning multi-class SVM advanced pattern recognizer which includes WDS historical data to create recognizable signature of different type and scale of leaks and breaks throughout the water distribution pipe line networks system, without implementation of the actual DMA.

4. Virtual DMA Model Formulation Using Mult-Class Support Vector Machine (SVM)

The application of computer aided systems, and multi-parameter monitoring sensors such as AMR and SCADA in water distribution pipeline networks to monitor the day to day operational activities allows billions of different data to be gathered, analyzed, which enables the WUC to act in the shortest amount of time. This results in creating a significant opportunity for water utility companies to search for innovative approaches for early leak detection and classification models. One of the advanced statistical approaches that could be used for leaks detection and classifications is the use of multi-class SVM advanced pattern recognizer [9, 10, 17, 19, 27, 33, 35, 41].

Many researchers have indicated that the SVM model has been proven very effective in detecting small leaks which other system could not be able detect using traditional method. The SVM model also allows avoiding the large sample requirements for anomaly classification [9, 10, 17, 19].

5. SVM & KSVM Model Formulations and Application Approach

The first theory and algorithm about SVM were originally established by Vapnik, V. N. [42, 43, 46], since then have been applied to solve many practical problems since 1990s. SVM has two major benefits like maximizing the margin and the kernel trick. This section reviews the support vector machine methodology in pattern recognition and classification. We choose SVM as our basic classier because SVM has been proven very effective in many research results, and are able to deal with large dimensions of feature space [4, 5, 17]. Consider binary classification task in which we have a set of training patterns. In put space, and let $x \in \mathbb{R}^d$ be the input space where \mathbb{R}^d is the d-dimensional Euclidean assigned one of two classes, w_1 and w_2 with and let $y_i = -1, 1$ be the output space denoted a binary

(normal/abnormal) decision (normal = no leaks exist, abnormal = leaks exist. Denote the linear discriminant function the entire equation is cited in [5, 17, 42, 43, 46]. With decision rule

$$g(x) = w^T x + w_0 \tag{1}$$

$$g(x) = w^T x + w_0 \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow X \in \begin{cases} w_1 = \text{with corresponding numeric value } y_i = +1 \\ w_2 = \text{with corresponding numeric value } y_i = -1 \end{cases}$$

Thus, all training points are correctly classified

$$\text{if } \begin{cases} x_i * w + b \geq +1 \text{ for } y_i = +1 \\ x_i * w + b \leq -1 \text{ for } y_i = -1 \end{cases}$$

If the water utility establish a threshold, as indicated on Table 1, and if the system pattern passes the threshold normal, if not it will be signal abnormal operation of the system

$$\text{Normal if } \sum_{i=1}^d w_i x_i > \text{water utility threshold}$$

$$\text{Abnormal if } \sum_{i=1}^d w_i x_i < \text{water utility threshold}$$

6. SVM Kernel Functions

There are different kinds of kernel functions that have been used for the SVM in finding the optimal solution. Such as such as the linear, polynomial kernel, sigmoidal kernel and the most popular one, radial basis function kernel, out of theses function, the polynomial kernel, sigmoid kernel and radial basis kernel function (RBF) are the most frequently used functions. RBF has fewer parameters than a polynomial kernel and used most often in general cases, because of its ability for better classification results [26, 32, 34, 35]. Consequently, the RBF is an effective choice for the kernel function [38]. Therefore, this study employs an RBF kernel function in the SVM to discover the optimal solution. The RBF kernel function expressed as:

$$K(x_i, x_j) = \exp[-\gamma \|x_i - x_j\|^2] \tag{2}$$

where: $-\gamma$ represents a parameter inversely proportional to the width of the Gaussian kernel.

7. Model Formulation Approach & Development of Multi-class-SVM Classifier

The research model formulation proposed for this study includes the support vector machine (SVM) multiple classifiers approach which have different steps that includes designing of input data and associated vector, trash hold formulation and classification principle etc. a single SVM only resolves two-classifier problems that is (+) or (-), since we have different scenarios, we set up the multiclass-SVM model composed by several classifiers which can identify different proposed signature of the pipeline networks leaks as described below. The

proposed multi-class SVM is based on recursively dividing the different signature of the pipeline networks leaks in to two disjoint groups that will decide in which of the groups the incoming unknown data from the WS pipeline system should be assigned. The recognition of each of the pipeline networks leaks signature starts according to assigned six different states of operational activities namely Normal, Abnormal (6-15%), Burst ($\geq 15\%$), High (15%), Medium (15%), and Low (6-8%) leaks from the network into consideration for virtual DMA. Table 1 below shows World Bank and IWA suggested leaks from WDS network, and used in this research for thresholds/signature formation of leak detection & classification.

Table 1: World Bank and IWA suggested Max and Min Leaks from WDS Network.

Normal	0-5%
Abnormal	6-15%
Break	>15%
High	15%
Medium	9%
Low	6-8%

Each node of the support vector machine (SVM) identifies pipeline leaks signature made based on assigned input pattern and the upcoming universal data from the sensor refer Figure 1. Each of these data has sub-groups that may contain multiple leaks signatures; therefore for the i^{th} SVM, the anomaly recognition output of i^{th} pipeline state will be (+1) while the outputs of the other five signatures of leaks will be (-1). The six leaks signature model formation and codes for Normal, Abnormal, Break or Burst, High leak, Medium leak and Low leak., and the output variables of multiple classifiers are (+1, 1, 1, 1, 1,1), (1, +1, 1, 1, 1, 1,1), (1, 1, +1, 1, 1, 1,1), (1, 1, 1, +1, 1, 1,1), (1, 1, 1, 1, +1, 1,1), and (1, 1, 1, 1, 1, +1) respectively. Figure 1 shows the general system architecture process of the proposed system. For Additional calibration the outputs of SVM classifiers conditional posterior probability have been used to interpret them as probability estimates.

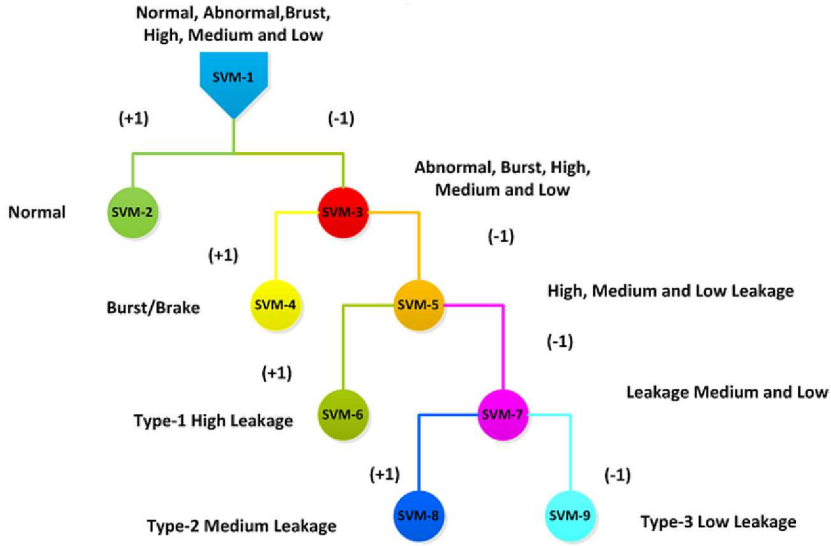


Fig. 1: Illustrate Of Multi-Class SVM Leak Detector & Classifier Model Formulation.

8. WS Pipeline Leaks Signature Identification

The WS pipeline leaks signature identification using multiclass SVM is based on equation 3 and 4 described below [26, 32, 34, 35].

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \right) \quad (3)$$

where:

b is threshold value of classification,

$K(x_i, x_j)$ is the kernel function,

x_i is training sample,

y_i is measured sample

$Y = \{-1, +1\}$ is the classification level,

α_i is the Lagrange coefficient vector satisfies

$$0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^n \alpha_i y_i = 0, \quad i = 1, 2, 3, \dots, n$$

where: C is the penalty factor.

Therefore, the RBF kernel function can be expressed as [26, 32, 34, 35].

$$K(x_i, x_j) = \exp [-\gamma \|x_i - x_j\|^2] \quad (4)$$

where: C and γ are parameters. The γ is inversely proportional to the width of the Gaussian kernel.

Due to the different combinations of parameters values C and γ are the most important factors that affect directly the outcome precision. During the modeling process the properties correspond to C and γ different. Therefore, maximum learning precision will occur in the vicinity of the straight line for by using and as the co-ordinates of parameter space and parameter combination for C and γ [26, 32, 34, 35].

9. Model Calibration and Validation

9.1 Mapping the SVM Outputs into Posterior Probabilities

The proposed multi-class SVM advanced pattern recognizer approach automatically induces classifier of different normal and abnormal operational pattern, that makes it possible to determine which type of leaks are more likely belong to a certain class (+1) or (-1). However, due to the complexity of the water distribution pipeline network and the diversity of the pipeline infrastructure (size, edge, material, etc.) depending only on the outcome of SVM classification is not reliable if the classification decision is cost sensitive.

Therefore, it is recommended to convert the outputs of SVM into well calibrated posterior probabilities and fitted it to sigmoidal function.

This section will describe the method we used to map the SVM output to posterior probability. The output of the SVM is not probabilistic. However, J. Platt [14, 31] proposed a simple solution to map the SVM outputs into posterior probabilities to compute the possible estimated output of the SVM class given the output of $p(y = +1/f(x))$ and $p(y = -1/f(x))$ of $f(x)$, using Bayes' rule probability, and assuming GMM of equal variance and sigmoid function. Using posterior probabilities method is by so far the most popular and

common method to transform uncelebrated SVM outputs, and this transformation can be carried out using equation (5) and (6) below [12, 29, 44, 45].

$$p(y = 1/f(x)) = \frac{1}{1 + \exp(A_j f(x) + H)} \quad (5)$$

where A_j and H can be determined by minimizing a negative log-likelihood function and $f(x)$ is the decision value of training data from the following equation

$$\min - \sum_{k=1}^n \left(t_k \log \left(p(y_k = \frac{1}{x_k}) \right) + (1 - t_k) \log \left(1 - p(y_k = \frac{1}{x_k}) \right) \right) \quad (6)$$

where $t_k = (y_k + 1)/2$ denotes the probability target and y_k is the classification label of sample i .

10. Preparing Training Data

The multi-class SVM proposed for this research is trained on six different signatures of leaks in WDS pipeline networks under consideration. Since it is very difficult to get the training sets from actual leaks, we used EPANET hydraulic modelling to simulate leaks scenario data using the EPANET emitter function which can simulate different size of pipeline holes, and then these generated values are used to train the SVMs on a number of different signatures of leaks in the pipe network, and used to detect and classify leaks [6, 10, 32]. The leaks detection algorithms developed are based on the assumption that all pipeline networks at Lille University “Zone-6” research facility are undamaged with multiple matrices of dependent and independent parameters. Let x_i be a $1 \times n$ row vector containing measurements that represent the operational state of Lille University “Zone-6” research area at a given time and this vector belongs to the input space \mathcal{R}^d .

$$x_i = [x_{i1} \quad x_{i2}, \quad \dots, \quad x_{in}]$$

This training set consisting of m data points, which is denoted by a $m \times n$ matrix representing a certain behavior of the model node matrix X :

$$x_i = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ \vdots & \ddots & & & \\ x_{i1} & & x_{ij} & & \\ \vdots & & & \ddots & \\ x_{m1} & & & & x_{mn} \end{bmatrix} \quad (7)$$

Lille University “Zone-6” water distribution networks are pressurized pipe systems which are subjected to a wide range of operational and loading conditions that vary with time. This operational state of water distribution pipeline network systems and the simulation of the hydraulic behavior of the system, for the given network characteristics and demands depends on dependent and independent parameters. Such as pressure, discharge or flow rate, pipe diameters, pipe lengths, and head loss coefficients. Therefore to simulate the hydraulic behavior of a system for the given network characteristics and demands which change with time to time and to get training data, let assume we have row data containing

measurements of independent parameters that represent the diagonal matrix of the pipeline network model nodal behavior and represented by ($HM_{Indp.Para}$) as

$$HM_{Indp.Para} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ \vdots & \ddots & & & \\ x_{i1} & & x_{ji} & x_{jn} & \\ \vdots & & & \ddots & \\ x_{m1} & & x_{mi} & & x_{mn} \end{bmatrix} \quad (8)$$

where i represents different kind of the independent parameters such as the roughness coefficients (C), nodal demands (ND), reservoir or tank water levels (WL), pipe diameters (D), pipe lengths (L), head loss coefficients (H), and so forth and j is the number of each indirect parameter (i). For example, if i is the roughness coefficients then j is the number of pipes, and if i is the nodal demands, then j is the number of nodes. Let also we have the simulation of a calibrated hydraulic model data set of dependent parameters, ($HM_{Dep.Para}$) as:

$$HM_{Dep.Para} = \begin{bmatrix} y_{11} & \dots & y_{i1} & \dots & x_{in} \\ \vdots & \ddots & & & \\ y_{j1} & & y_{ji} & & y_{jn} \\ \vdots & & & \ddots & \\ y_{m1} & & y_{mj} & & y_{mn} \end{bmatrix} \quad (9)$$

where i is the number of dependent parameters such as pressure (P), head discharge (Q), flow velocity (V), and j is the number of nodes. The relation between the independent parameter and dependent parameter (pressure (P), head discharge (Q), flow velocity (V)), with pressure can be express as follows using Todini from equations (10)-(12) [18, 20, 39, 40].

$$\begin{bmatrix} A_{pp} & \dots & A_{pn} \\ \vdots & \ddots & \vdots \\ A_{np} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} Q \\ \vdots \\ H \end{bmatrix} = \begin{bmatrix} -A_{p0} & H_0 \\ \vdots \\ -q \end{bmatrix} \quad (10)$$

where:

$Q^T = [Q_1, Q_2, \dots, Q_{n_P}]^T$ is the $[1, n_{PP}]$ unknown pipe discharge (Q),
 $H^T = [H_1, H_2, \dots, H_{n_P}]^T$ is the $[1, n_{PP}]$ unknown nodal head (NH),
 $H_0^T = [H_{n_{n-1}}, H_{n_{n-2}}, \dots, H_{n_t}]^T$ is the $[1, n_t - n_n]$ known pipe nodal head,
 $q^T = [q_1, q_2, \dots, q_{n_n}]^T$ is the $[1, n_n]$ known pipe nodal demand,

where:

n_p is the number of pipes,
 n_n is the number of nodes,
 n_t is the total number of nodes in the network,
 $n_t - n_n$ is the numbers of the nodes with known head.

In equation (10) A_{pp} is the diagonal matrix which elements, include minor losses, and defined for $K \in 1, n_p; i \in 1, n_i; j \in 1, n_t$ as

$$A_{pp}(k, k) = r|Q_{ij}|^{n-1} + m|Q_{ij}| \quad (11)$$

For pipes

$$A_{pp}(k, k) = -w^2 \left(h_0 - r \left(\frac{Q_{ij}}{\omega} \right) \right) / Q_{ij} \quad (12)$$

These entire coefficient $r, m, n\omega, a_0, b_0, c_0$ are relevant to the specific pipes, and the actual network topology is described by means of the topological incidence matrix $A_{pp} = [A_{pn}/A_{p0}]$ defined as

$$A_{pp}(i, j) = \begin{cases} -1 & \text{- if the pipe } j \text{ leaves node } i \\ 0 & \text{- if the pipe } j \text{ is not connected to node } i \\ +1 & \text{- if the flow of pipe } j \text{ enters node } i \end{cases}$$

Leaks flow rates in equation (19) can be calculated by using A_{pp} as a diagonal scalar product matrix (i.e. element by element product) if the water demand node that having significant is not caused by hydraulic head, or a nonlinear function of the pressure, and the actual head H_i and the terrain elevation Z_i , using the E.Todini formulation, and can be express this function as [39, 40]:

$$A_{pp}(i, j) = \begin{cases} -1 & H_i \leq Z_i \\ \frac{q_i}{H_i} \left(\frac{H_i - Z_i}{H_i^* - Z_i} \right) & Z_i \leq H_i \leq H_i^* \\ +1 & H_i \leq H_i^* \end{cases} \quad (13)$$

where H_i^* is the required nodal head.

Assuming leaks q_{k-Leak} along pipe k , the background leaks model can be expressed as in equations (14). For $q_{i-act}(p_i)$ the following relationship will be used here [18].

$$P_{i-act} = \begin{cases} q_{i-design} & \text{for } P_i \leq P_{i-ser} \\ q_{i-design} \left(\frac{P_i - P_{i-min}}{P_{i-ser} - P_{i-min}} \right)^{1/2} & \text{for } P_{i-min} \leq P_i \leq P_{i-ser} \\ 0 & \text{for } P_i \leq P_{i-min} \end{cases} \quad (14)$$

where:

P_{i-ser} is design operational pressure, used for network design purposes

P_{i-min} is the intermediate operating pressure,

P_{i-act} is the actual demand

i index for nodal-level variables

K index for for pipe-level variables.

For demands that are not pressure-driven, Equations (14) become $P_{i-act} = P_{i-design}$.

If we assume leaks P_{k-Leak} along pipe k , the background leaks model can be expressed using the formula [18]:

$$q_{k-leak} = \begin{cases} \beta_K l_K (P)^{\alpha_K} & \text{for } If P_K > 0 \\ 0 & \text{for } If P_K \leq 0 \end{cases} \quad (15)$$

where: P_k is the average pressure in the pipe computed as the mean of the pressure values at the end nodes i and j of the k^{th} pipe, and l_k is the length of that pipe, α_k, β_k variables

denote the two leaks model parameters, P^{pipes} is average pressure vector can be computed as

$$P^{pipes} = \frac{(|\overline{A_{pn}}| [P^{nodes}/P_0^{nodes}])}{2} \quad (16)$$

where:

P^{nodes} is the pressure vector of unknown nodal heads

P_0^{nodes} is the pressure vector of known nodal heads

$|\overline{A_{pn}}|$ is the absolute value of the topological matrix.

For this research purpose the operational pressure of each facility in “Zone-6” connection and the water main nodes have been taken for leaks allocations based on connection and nodal based analysis. Using equation (17), and by introducing the emitter specifications, through the discharge coefficient C_{i-node} at model network nodes, L_{i-node} leaks can be express as [6, 10, 18]:

$$Leakage_{(i-node)} = C_{(i-node)} \otimes (P_{(i-node)})^{N1} \quad (17)$$

where $N1$ is the chosen pressure coefficient and $C_{(i-node)}$ is the nodal leaks discharge or emitter coefficient. The equation used to calculate the $N1$ value based on the changes in average zone pressure (AZP) pressure and physical losses in the network is as follows:

$$N1 = \frac{\log(L_1/L_0)}{\log(P_1/P_0)} \quad (18)$$

The EPANET emitter coefficient has a unit of flow rate per unit pressure and has to specify for selected pipeline networks for the entire network, and can be range from 0.001 to 0.00075 [6, 10].

11. Leaks Detection Algorithms (LDA) and Classification Code Matrix

The leaks detection algorithms (LDA) and classification code matrix for the proposed multiclass SVM the following the kernel function and the regularization parameter for training phase used. For this purpose the binary SVM classification the code matrix of multi-class SVM leak detector and classifier where the pipeline network operational states are as Normal (NR), Abnormal (AB), Break or Burst (BR), High leak (HL), Medium leak (ML) and Low leak (LL) also presented below in Table 2.

Table 2: Code matrix

NR-1	+1	-1	-1	-1	-1	-1
AB-2	-1	+1	-1	-1	-1	-1
BR-3	-1	-1	+1	-1	-1	-1
LL-4	-1	-1	-1	+1	-1	-1
ML-5	-1	-1	-1	-1	+1	-1
HL-6	-1	-1	-1	-1	-1	+1

12. Experimental- Case Study Lille University

In this section, we present the results of the experiment using multi-class SVM advanced pattern recognizer explained above, for the research facilities of Lille University “Zone-6” study area.

12.0.1 Introduction

The University Lille 1 was founded in 1854 in Lille, France, and ranked as one of the world top 200 universities. The university of Lille operates European Community Sponsored Smart Urban Network Center with Critical Infrastructure Operators Center (Water, energy etc.), and has been Selected EC Demonstration site for smart water 10 million euro. Lille University water pipeline network systems are divided into different supply zones Z1, Z2 etc. This research is carried out for the “Zone-6” project area.

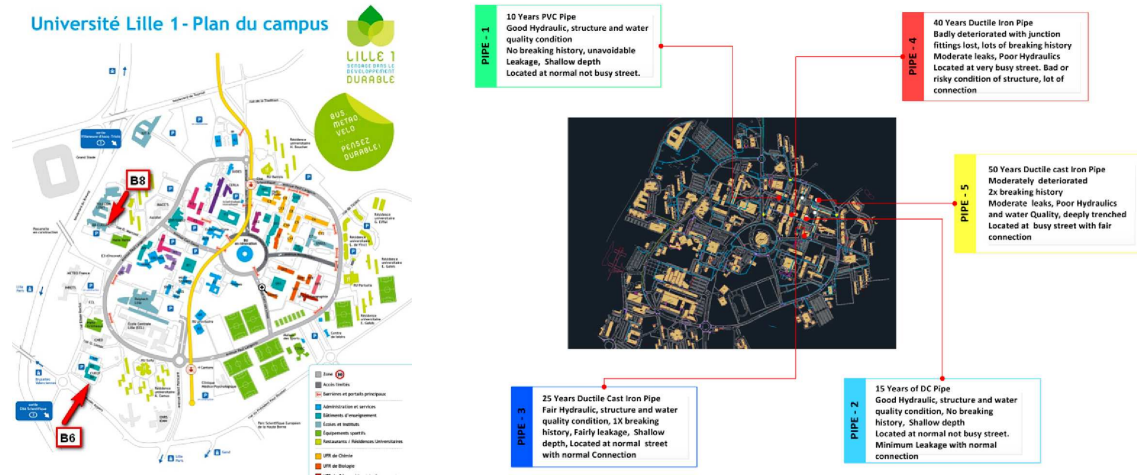


Fig. 2: General Location of the Lille University Campus and “Zone-6” Research Area Pipeline Mains & Existing Condition.

12.1 Lille Universitys Water Pipeline Networks

Lille University water pipeline network systems are divided into different supply zones Z1, Z2, ... etc. This case study carried out for the so called “Zone-6” project area. Currently, there are approximately 3.58 km of water pipelines within the “Zone-6” with diameter of 150mm to 300mm. The aged of the pipe lines ranges from 10-50 years with different materials such as cast iron, ductile iron, and PVC. This network has operating pressure of approximately 4- 5 bars or 58-72 Psi [36, 37]. The Pipelines data from the university database are used for parameters and criteria formulation required to develop multi-class SVM approach. These include structural data for the pipes (e.g. diameter, length of pipe, material, laying year, and soil conditions, co-ordinates, joint type, ... etc.).

13. Model Demonstration

The selected research area the so called “Zone-6” of Lille University has 58 pipes with total length of 3.85 km. connected by 60 nodes and supplied by gravity from one elevated tank with a total head of approximately 120 m. The length of the pipe varies from 20m to 400m and the diameters of the pipes varies from 150mm to 700mm. base demands at different node varies from 40l/s to 90l/s and demand multipliers ranges from 0.38 at 5.0am to 1.49 at 9am. For this research the WDS has been considered undamaged system with no breaks or considerable leaks, and all the parameter values are considered as calibrated values in normal conditions. We also assume a minimum of pressure 20 pounds per square inch (psi) at all water taps including fire hydrant locations under all conditions of design flow will be maintained [36, 37].

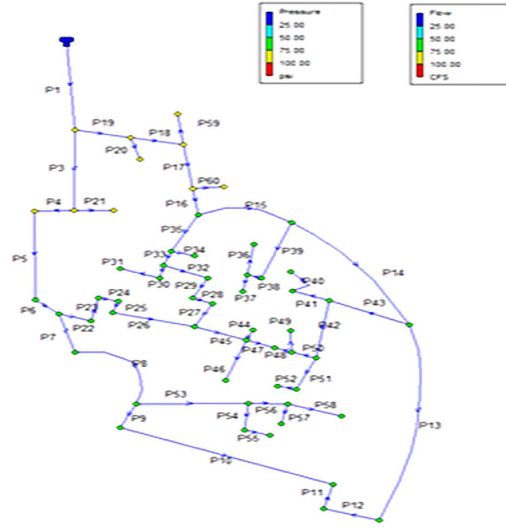


Fig. 3: EPA-Net Layout of the Research Area 11Lille University Zone-6” and the EPANET Simulation Set up for Different Leaks and Breaks Scenario (Typ.) The P Numbers Indicates the Number of Nodes WDS. Note that Pipe Layout is Not in Scale. The Results of leaks detection and classification after calibration and validation presented below.

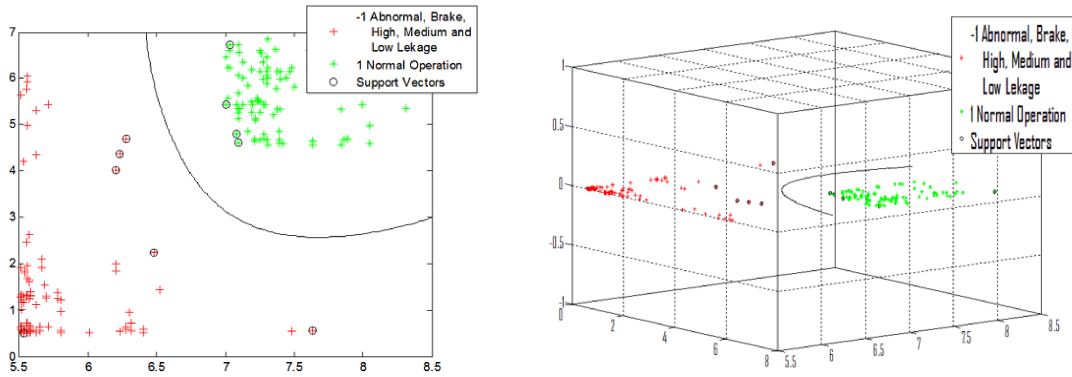


Fig. 4: 2D (Left) and 3D (Right) Results of Trained KSVM Plot for Normal in Green and Abnormal in Red Data. The Support Vectors are shown as “O”

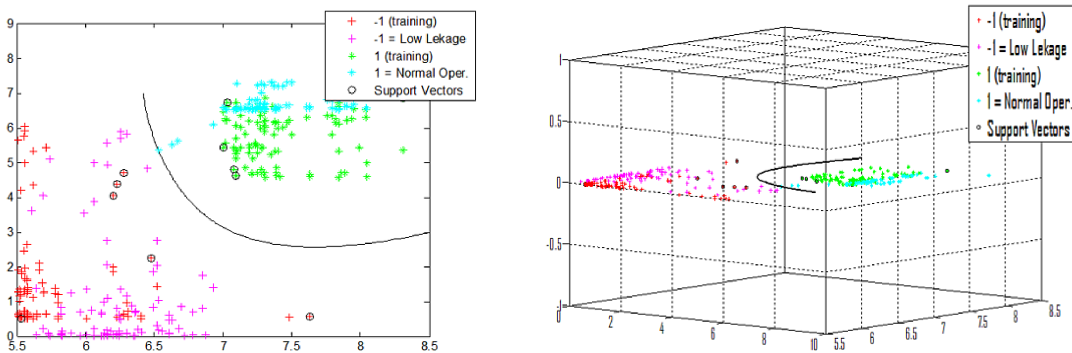


Fig. 5: 2D (Left) and 3D (Right) Results of KSVM Plot for Trained Normal (green) , Abnormal (red) and the Predicted or Classified Low Level or Class-1 Leaks (magenta) and Normal Operation pattern (Cyan). The Support Vectors are Shown as “O”.

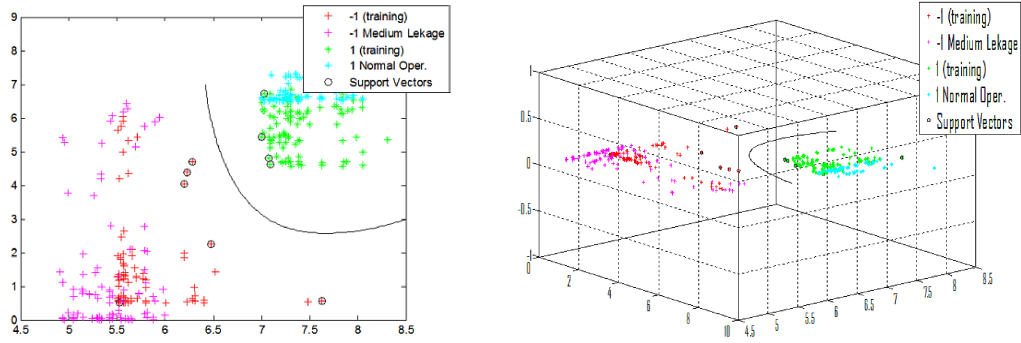


Fig. 6: 2D (Left) and 3D (Right) Results of KSVM Plot for Trained Normal (green) , Abnormal (red) and the Predicted or Classified Medium level or Class-2 leaks (magenta) and Normal Operation pattern (Cyan). The Support Vectors are Shown as “O” .

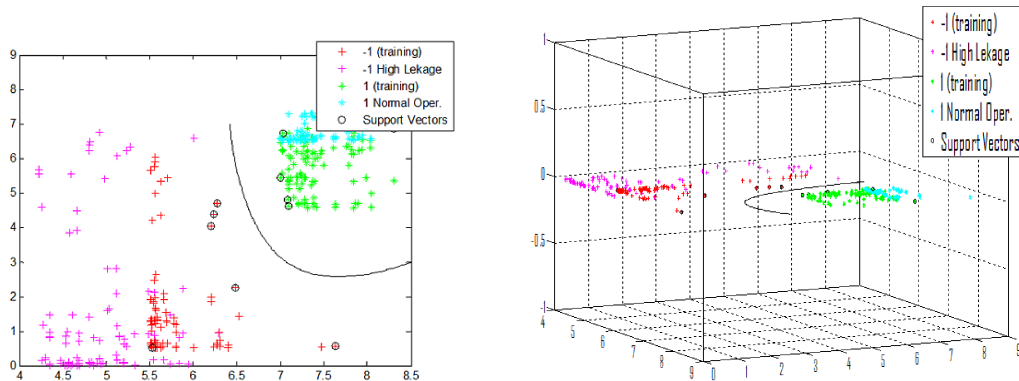


Fig. 7: 2D (Left) and 3D (Right) Results of KSVM Plot for Trained Normal (green), Abnormal (red) and the Predicted or Classified High Level or Class-1 Leaks (magenta) and Normal Operation pattern (Cyan). The Support Vectors are Shown as “O” .

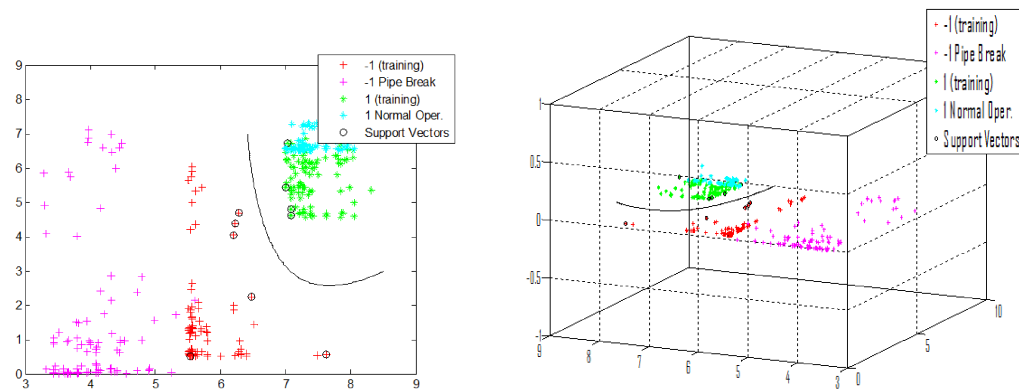


Fig. 8: 2D (Left) and 3D (Right) results of KSVM Plot for Trained Normal (green), Abnormal (red) and the Predicted or Classified as Break or Burst or Class-4 leaks.

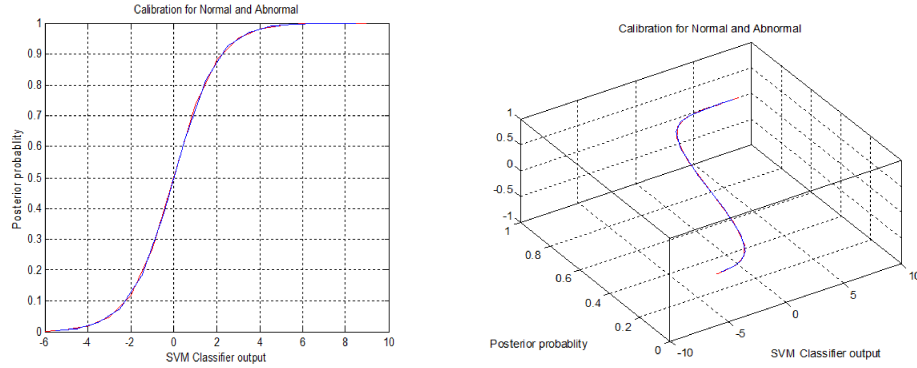


Fig. 9: Plot of 2D (Left) and 3D (Right) Calibrated and Validated SVM Outputs Posterior Probabilities Fitted on Sigmoid Function for Lille University water Distribution Network Zone-6 Normal and Abnormal Training Data Set. The Abscissa is the Classified Score (SVM Distance), and the Ordinate is the Calibrated Posterior Probability Produced by Sigmoid Function.

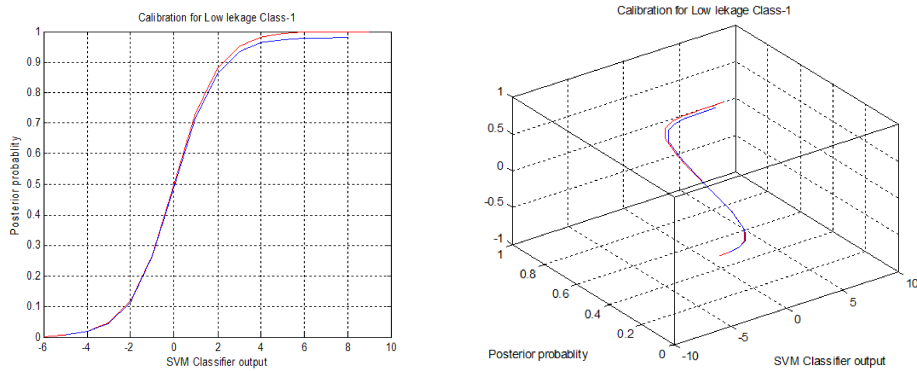


Fig. 10: Plot of 2D (Left) and 3D (Right) Calibrated and Validated SVM Outputs Posterior Probabilities Fitted on Sigmoid Function for Lille University Water Distribution Network Zone-6 Low Leaks Class-1 Data Set. The Abscissa is the Classified Score (SVM Distance), and the Ordinate is the Calibrated Posterior Probability Produced by Sigmoid Function.

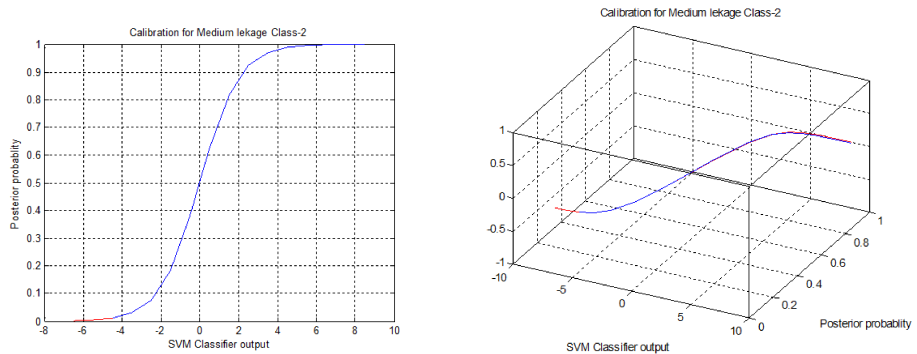


Fig. 11: Plot of 2D (Left) and 3D (Right) Calibrated and Validated SVM Outputs Posterior Probabilities Fitted on Sigmoid Function for Lille University Water Distribution Network Zone-6 Medium Leaks Class-2 Data Set. The Abscissa is the Classified Score (SVM Distance), and the Ordinate is the Calibrated Posterior Probability Produced by Sigmoid Function.

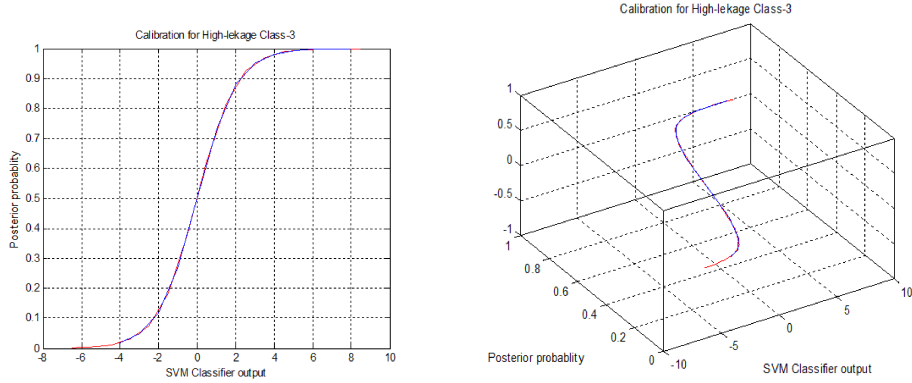


Fig. 12: Plot of 2D (Left) and 3D (Right) Calibrated and Validated SVM Outputs Posterior Probabilities Fitted on Sigmoid Function for Lille University Water Distribution Network Zone-6 High leaks Class-3 Data Set. The Abscissa is the Classified Score (SVM Distance), and the Ordinate is the Calibrated Posterior Probability Produced by Sigmoid Function.

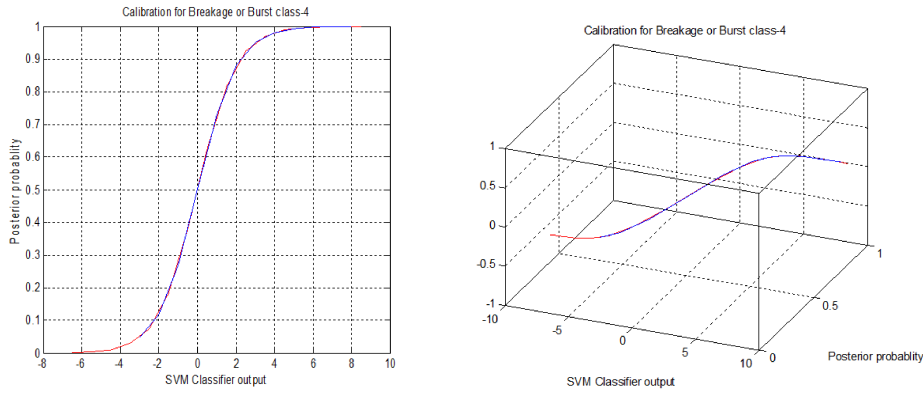


Fig. 13: Plot of The 2D (Left) and 3D (Right) Calibrated and Validated SVM Outputs Posterior Probabilities Fitted on Sigmoid Function for Lille University Water Distribution Network Zone-6 Breaks or Burst Class-4 Data Set. The Abscissa is the Classified Score (SVM Distance), and the Ordinate is the Calibrated Posterior Probability Produced by Sigmoid Function.

14. Discussion and Findings

In this research paper we investigated and analyzed the concept of virtual DMA municipal water distribution pipelines leaks detection and classification approach using multi-class SVM advanced pattern recognizer with the strategic integration of the result to asset management.

- First, the application of SVMs machine learning techniques used for this research results and analysis demonstrated promising performance, which leads us to conclude that virtual DMA Multi-class SVM Advanced Pattern Recognizer could be successfully employed for leaks detection and classification in water distribution system.
- Second, the leaks scenario dataset generated to represents the different signature of leaks in the WS pipeline network for this research were through EPANET hydraulic model, and analyzed using multi-class SVM advanced pattern recognizer. The proposed model results have shown the potential possible future application of virtual

DMA. However, the model leaks signature limit could be maximized by using quality data gathered with advanced multi-parameter monitoring sensors.

- Third, the other interesting finding is, if the constraints are not known or can change over time when the SVM model is constructed, we need a method which is readily capable of adapting and responding to the current resource constraints. Therefore, the use of classifier conditional posterior probabilities as calibration and decision function for classification instead of using directly SVM output function can increase classification performance confidences.

15. Conclusion

This research paper modelled and simulated the concept of virtual DMA leaks monitoring and classification system using multi-class support vector machine (SVM) advanced pattern recognizer. It attempts to demonstrate the applicability of Virtual DMA for early leaks detection and monitoring. The general approaches and the overall analysis and result shows good promise for the applications of this model for the benefit of system operators and decision makers of water utility companies for selection of which pipeline infrastructure required urgent action, and engineer the optimal alternative of rehabilitation and replacement (R&R) maintenance strategies and leaks monitoring and classification using virtual DMA. Furthermore, this research approaches also facilitates for water utility companies which are searching for innovative technology for early leaks detection and monitoring system for better managing their WDS pipeline networks.

References

- [1] J. Almandoz, E. Cabrera, F. Arregui, E. Cabrera Jr, & R. Cobacho. Leaks assessment through water distribution network simulation. *Journal of water resources planning and management*, 131(6), 458-466, 2005.
- [2] *American Water works Association (AWWA, 1999) and by the IWA (Alegre et al., 2000).*
- [3] *American Water Works Association. Water audits and leak detection* Vol. 36, 1999.
- [4] R. Bi, Y. Zhou, F. Lu & W. Wang, Predicting Gene Ontology functions based on support vector machines and statistical significance estimation. *Neuron-computing*, 70(4), 2007.
- [5] C. J. Burges, A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), pp. 121-167, 1998.
- [6] A. M. Cassa, J. E. Van Zyl & R. F. Laubscher. A numerical investigation into the effect of pressure on holes and cracks in water supply pipes. *Urban Water Journal*, 7(2), pp. 109-120, 2010.
- [7] L.C. Cheong Unaccounted for Water & the Economics of Leak Detection, *Proceedings of the 18th International Water Supply Congress and Exhibition*, 15-31 May 1991, Copenhagen, published in *Water Supply*, 9:3&4:IR1.1, 1991.
- [8] T. Chis, Pipeline Leak Detection Techniques. *arXiv preprint arXiv:0903.4283*, 2009.
- [9] J. Cui, Y. Wang, A novel approach of analog fault classification using a SVM, 2010
- [10] D. De Silva, J. Mashford, & S. Burn, Computer aided leak location and sizing in pipe networks. *Urban Water Security Research Alliance*, 2009.
- [11] D.M. Delgado, Infrastructure Leaks Index (ILI) as a Regulatory and Provider Tool. *Arizona: Department of Civil Engineering & Engineering Mechanics*, 2008.
- [12] P. A. Devijver, & J. Kittler, Pattern recognition: A statistical approach. *Englewood Cliffs, NJ: Prentice/Hall International*, p. 448, 1982.
- [13] Drinking Water Protection Division of the EPA. *A Review of Control and Mitigation of Drinking Water Losses in Distribution Systems*, Apx.B-3. EPA MC-4606M EPA 816-D-09-001, 2009.
- [14] J. Drish, Obtaining calibrated probability estimates from support vector machines. *In Final Project for CSE 254: Seminar on Learning Algorithms*, June 2001.

- [15] M. Farley, Are there alternatives to the DMA?. *Asian Water*, pp. 10-16, 2010.
- [16] R. Fletcher, Smart ball A new approach in pipeline leak detection. *Proceedings Int. Pipeline Conference IPC2008*, Calgary, Canada, 2008 Sept/Oct.
- [17] M. G. Genton, Classes of kernels for machine learning: a statistics perspective. *The Journal of Machine Learning Research*, Vol. 2, pp. 299-312, 2002.
- [18] G. Germanopoulos & P. W. Jowitt, Leaks reduction by excess pressure minimization in a water supply network. *In ICE Proceedings* Vol. 87, No. 2, pp. 195-214). *Ice Virtual Library*, 1985, 1989, June.
- [19] M. Girolami, Mercer kernel-based clustering in feature space. *Neural Networks*, IEEE Transactions on, 13(3), pp. 780-784, 2002.
- [20] O. Hunaidi, Detecting leaks in water distribution pipes. *Institute for Research in Construction, National Research Council of Canada*, 2000
- [21] A. K. Jain & R. C. Dubes, Algorithms for clustering data. *Prentice-Hall Inc*, 1988 and L. Kaufman & P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis. *Wiley. Com*, Vol. 344, 2009.
- [22] M. John, Water21 by the IWA *Water Loss Task Force*, 2004.
- [23] E. Kober & G. Gangl, New Monitoring Methodology for Water Distribution Systems. *In IWA Conference Water Loss*, 2009.
- [24] D.W. Kurtz, Developments in a Free-Swimming Acoustic Leak Detection System For Water Transmission Pipelines. *Precede Pipelines 2006*, ASCE, July 30 - Aug 2, Chicago, 2006
- [25] A. O. Lambert, International report: water losses management and techniques. *Water Science & Technology: Water Supply*, 2(4), pp. 1-20, 2002.
- [26] D. Misinas, Failure Monitoring and Asset Condition Assessment in Water Supply Systems. *Vilniaus Gedimino technikos universitetas*, 2008.
- [27] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, & B. Scholkopf, An introduction to kernel-based learning algorithms. *Neural Networks, IEEE Transactions on*, 12(2), pp. 181-201, 2001.
- [28] News on Lille 1's webpage, <http://www.univ-lille1.fr/Accueil/Actualites?id=26313>
- [29] R. Pilcher, Leak detection practices and techniques: a practical approach, *Water 21* December 2003 Issue 5.6 pp. 44-45, IWA Publishing, 2010.
- [30] R. Pilcher, S. Hamilton, H. Chapman, D. Field, B. Ristovski & S. Stapely, Leak Location and Repair, *IWA(IWA), Specialist Group on Efficient Operation and Management of Urban Water Distribution Systems, Water Loss Task Force*, 2007.
- [31] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3), 61-74, 1999.
- [32] R. Puust, Z. Kapelan, D. A. Savic & T. Koppel, A review of methods for leaks management in pipe networks. *Urban Water Journal*, 7(1), pp. 25-45, 2010.
- [33] R. Sangeetha & B. Kalpana, Performance Evaluation of Kernels in Multiclass Support Vector Machines. *Training*, 2, 2, 2011.
- [34] B. Schlkopf, C. J. Burges & A. J. Smola (Eds.), 1999. Advances in kernel methods: support vector learning. *The MIT press*. M. G. Genton, 2002.
- [35] A. J. Smola & B. Schlkopf, A tutorial on support vector regression. *Statistics and computing*, 14(3), pp. 199-222, 2004.
- [36] T. Mamo, I. Juran, I. Shahrou, Prioritization of Municipal Water Mains Leakss for the Selection of R&R Maintenance Strategies Using Risk Based Multi-Criteria FAHP Model. *Journal of Water Resource and Hydraulic Engineering* Vol. 4, 2013.
- [37] T. Mamo, I. Juran, I. Shahrou, Urban Water Demand Forecasting Using the Stochastic Nature of Short Term Historical Water Demand and supply Pattern. *Journal of Water Resource and Hydraulic Engineering*, Vol. 3, 2013.
- [38] X. Tang, L. Zhuang, J. Cai & C. Li, Multi-fault classification based on support vector machine trained by chaos particle swarm optimization. *Knowledge-Based Systems*, 23(5), 2010.
- [39] E. Todini, A more realistic approach to the extended period simulation of water distribution networks. *Advances in water supply management*, 173-184, 2003.
- [40] E. Todini, Towards realistic extended period simulations (EPS) in looped pipe network. *In Proc., 8th Annual Int. Symp. on Water Distribution Systems Analysis*. Reston, August 2006.
- [41] S. Trow & M. Farley, Developing a strategy for leaks management in water distribution systems. *Water Science & Technology: Water Supply*, 4(3), pp. 149-168, 2004.

- [42] V. N. Vapnik, Statistical learning theory, 1998.
- [43] J. Wan, Y. Yu, Y. Wu, R. Feng & N. Yu, Hierarchical leak detection and localization method in natural gas pipeline monitoring sensor networks. *Sensors*, 12(1), pp. 189-214, 2011.
- [44] A. R. Webb, Statistical pattern recognition. *Wiley. com*, 2003.
- [45] L. Xu, Bayesian Ying Yang learning (I): a unified perspective for statistical modeling. *In Intelligent Technologies for Information Analysis. Springer Berlin Heidelberg*, pp. 615-659, 2004.
- [46] Y. Ying, Machine based theory for a classical open problem in pattern recognition. In Neural Networks *IEEE International Conference* Vol. 3, pp. 1546-1551, 1996.