Leadership in Sustainable Infrastructure

Leadership en Infrastructures Durables

Vancouver, Canada

May 31 – June 3, 2017/ *Mai 31 – Juin 3, 2017*

LEAK DETECTION MODEL FOR PRESSURIZED PIPELINES USING SUPPORT VECTOR MACHINES

Samer El-Zahab [1, 3], and Tarek Zayed [2, 4]

[1] Ph.D. Student, Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada H3G 1M8.

[2] Professor, Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada H3G 1M8.

[3] s_elzah@encs.concordia.ca

[4] zayed@encs.concordia.ca

ABSTRACT

Aging infrastructures, specifically pipelines, that were installed quite a while back and currently operating under poor conditions, are highly susceptible to the threat of leaks, which pose economic, health, and environmental threats. For example, in the year 2009, the state of Ontario lost 25% of its water supply solely due to leaks. The amount of lost water is equivalent to the volume of 131,000 Olympic swimming pools and worth 700 million Canadian dollars. Therefore, a need arises to develop an approach that allows condition monitoring and early intervention. This article proposes a model for a real-time monitoring system capable of identifying the existence of single event leaks in pressurized water pipelines. The model relies on wireless accelerometers placed within the network on the exterior of the pipelines. The vibration signal derived from each accelerometer was assessed and analyzed to identify the Monitoring Index (MI) at each sensor on the pipeline. The data collected from experimentation were analyzed by means of support vector machines (SVM) technique. A leak threshold was determined such that if the signal increased above the threshold, a leak status is identified. Experiments were performed on one inch cast iron pipelines, one inch and two inch PVC pipelines using single event leaks and the results were displayed. The developed models showed promising results with 98.25% accuracy in distinguishing between leak states and non-leak states.

Keywords: Leak Detection, Water Mains, Support Vector Machines, Accelerometers, Vibration Signals, Asset Management

1. INTRODUCTION

One billion individuals around the planet do not have access to clean drinking water (Krchnak 2016). Water distribution systems are highly susceptible to leaks, which often leads to potential water losses. Usually, 20% to 30% of the produced water is lost in transmission networks mainly due to leaks (Cheong 1991), and in some systems, this loss can surpass 50% of the produced water (AWWA 1987). The contribution of leaks to the total water loss within networks is estimated to be at 70% and is expected to rise in low maintenance locations (Van Zyl and Clayton 2007). Water loss is not the only outcome of leaks, as they also create problems at the social and environmental levels. As an example, the United Kingdom is estimated to open approximately 4 million holes into its road network to install pipes and repair water leaks. The overall cost of those damages is estimated to be 7 billion £ per year (10,073,140,000 US$) divided into 1.5 billion £ (2,158,305,000 US$) of direct damage costs, and 5.5 billion £ (7,914,500,000 US$) in social impact costs (Royal et al. 2011). Unattended leaks are susceptible to grow and thus allowing the introduction of pathogens and contaminants from the surrounding environment, which results in a major decrease in the quality of the supplied water and might cause harmful impacts on human life and other beneficiaries (Alkasseh et al. 2013).

The damages and negative impacts drew the attention of researchers to help develop a real-time monitoring system within water main networks that allows early detection of leaks and eventually optimally-timed repairs. Multiple models were developed to address the issue, nevertheless, several limitations were encountered on the levels of accuracy, device availability, applicability, false alarms, and impacts of external conditions. Hence, this research proposes a real-time monitoring system for pressurized water networks that is capable of detecting, localizing, and pinpointing leaks, by utilizing accelerometers, which are widely available in the market. Based on what preceded the objectives of this paper are:

(1) Present an overview of the technologies and approaches utilized in this paper.
(2) Conduct experiments on pressurized PVC and ductile iron pipelines using accelerometers.
(3) Develop a leak detection model using Support Vector Machines (SVMs) and vibration signals.
(4) Develop a leak classification model using SVM and vibration signal.

2. BACKGROUND

2.1 Leak Detection Phases

Hamilton (2009) divided the leak detection process into three main steps: Localizing, Locating, and Pinpointing. The first phase, localizing, is an approach by which the user aims to narrow down segments of a district metered area that is suspected to contain a leak or burst. In order to achieve this goal, multiple techniques are utilized, such as step testing, acoustic noise logging, all fitting surveys, mains fitting only, and district metered area

. The second phase, locating, is an approach that denotes a one-meter radius area within the previously identified segment as the leak area. One of the most prevalent techniques in this field is "Noise Correlation", which relies on assessing the speed of multiple sound signals at multiple locations and using the difference in detection time, thus identifying the leak location. The final phase, pinpointing, is an approach that aims at determining the exact location of the leak prior to excavation. The accuracy of the exact location is expected to be within a radius of 200 mm given that the equipment is used correctly by capable workmanship. Pinpointing techniques include mainly listening sticks, geophones, and hydrophones. A worker is expected to pass this device either within or above the expected leak location and therefore the location with the highest amplitude is expected to be the leak location (Hamilton 2009, El-Abbasy et al. 2016).

2.2 Accelerometers and Leak Detection

Approaches that utilize accelerometers only received more attention in recent years as a sole leak detection system using their capabilities to detect vibration signals that are emitted by leaks. Shinozuka et al. (2010), suggested the use of accelerometers in order to measure the variation in acceleration forces and record those variations to identify damaged areas and map the network. In terms of standalone vibrational models, Almeida et al. (2014) developed an improved technique for the interpretation and assessment of the data gathered from accelerometers, that are monitoring water pipelines, through a set of mathematical models. The model resulted in an improved identification of the frequency range of accelerometers, identification of time delays between two sensors, and the estimation of the wave speed of propagation. The derived values can be utilized by the correlation formula to pinpoint the leak's location as shown.

$$\boldsymbol{d_2 = \frac{d - cT_0}{2}} \qquad \textbf{(1)}$$

Where:

- c: wave speed
- d: the total distance between measurement positions
- d_2: distance from sensor 2 to the leak

- t_0: the time delay between sensor 1 and sensor 2

Additionally, on the level of vibration analysis, Martini et al. (2015) presented a model that relies solely on vibration signals within PVC pipeline. The model presented a vibration signal analysis approach that is capable of aiding in deciphering the signals and converting their values into a comprehensible index. The model proved to be highly capable of accurately identifying the slightest leaks within PVC pipelines. This model was further improved by Zahab et al. (2016) through further experimentation. The model utilized the previously developed index by Martini et al. (2015) in developing a regression model that is capable of pinpointing the leak location within PVC and ductile iron pipelines with an average accuracy of 95%.

On the level of accelerometers, Martini et al. (2015) proposed a model to utilize accelerometers for leak detection. Additionally, Martini suggested an approach for analyzing the signal received from the sensors. Their mathematical analysis approach can be summed in the following steps: (1) Determine acceleration reading per second in (g), (2) Each t = 100 second, the readings are collected and their standard deviation is determined, (3) After monitoring for several hours, the lowest 10 standard deviations are averaged to determine the lowest monitoring index as shown in Equation **(2)**.

$$MI_j = mean(\sigma_j, 10) \tag{2}$$

A value named Monitoring Index Efficiency (MIE) is determined by dividing the current monitoring index of any instant with the lowest monitoring index in no leak state of a given duration $MI_{j,\text{no leak}}$ as illustrated in Equation **(3)**. This equation allows the establishment of sensor specific values. For example, MI_0 can be unique for each sensor as well as the readings, therefore taking into consideration any pre-existing conditions and external factors. Martini originally utilized the maximum monitoring index of a duration, whereas, in this research, MIE is determined each t = 100 second.

$$MIE_x = \frac{MI_i}{MI_0} \tag{3}$$

Where:

- MIE: is the Monitoring Index Efficiency.
- x: can be either L or R representing left or right sensor respectively.
- MI_0: is the lowest recorded monitoring index at no leak state.
- MI_i: is the monitoring index of a given signal at time period i.

2.3 Support Vector Machines in Leak Detection

The support vector machines (SVM) algorithm is a supervised learning classification algorithm that was first proposed by Cortes and Vapnik (1995). The algorithm was developed to serve as an intelligent classification technique. One of the main original capabilities of SVM was to create decision-making thresholds and solution for binary classification problems. The SVM algorithm can be divided into two main categories: (1) Linear SVM for linearly separable problems, and (2) Non-linear SVM for problems that are not linearly separable and require much more complex solutions (Fletcher 2009).

The SVM algorithm has also been utilized to identify leak locations as proposed by Mashford et al. (2009). The article utilized pressure data and flow rate data to detect and identify the existence of leaks as well as leak location using the SVM algorithm as a classifier and as a regressor. The data was collected, using multiple experimentation procedures, via a hydraulic modeling system that is dubbed “EPANET”. The developed regression model via SVM algorithm had a high r-squared that is close to 100% as well as a mean square error close to zero. Yet, the model relies heavily on the quality of the data provided as well as the sensitivity of the sensors utilized. Furthermore, the SVM algorithm found further success in leak detection when coupled with other concepts such as the rough set theory as presented by Mandal et al. (2012). The rough set theory was used to create a set of rules that would exclude any predetermined and easily identifiable non-leak states or leak states to guarantee that the SVM algorithm would not be confused by any anomalies and then a trained SVM algorithm was used in order to classify the cases that were not ruled out by rough set theory. This approach had multiple advantages including (1) High accuracy, (2) Capability to replace mass balance and pressure point analysis, (3) Providing a detailed simulation of the pipeline understudy, and (4) Effective to utilize in old pipelines. The SVM algorithm is deemed to be a powerful supervised classification and regression tool that is capable of differentiating leak states, and aid with identifying leak location given that the data is well constructed.

3. METHODOLOGY

The leak detection model main goal is alerting the user of the existence of a leak in a certain segment instantly. Thus, to develop this model, the first task was to perform multiple experiments mainly on PVC and ductile iron pipelines of sizes one-inch and two-inch, as illustrated in Figure 1. Afterward, using the signal received from the accelerometers in simulated leak and no-leak, and their respective collected MIE, the model input was organized. The form of the model input is represented by the predetermined MIE and the relevant state, for example (MIE = 0.982, State = No Leak).

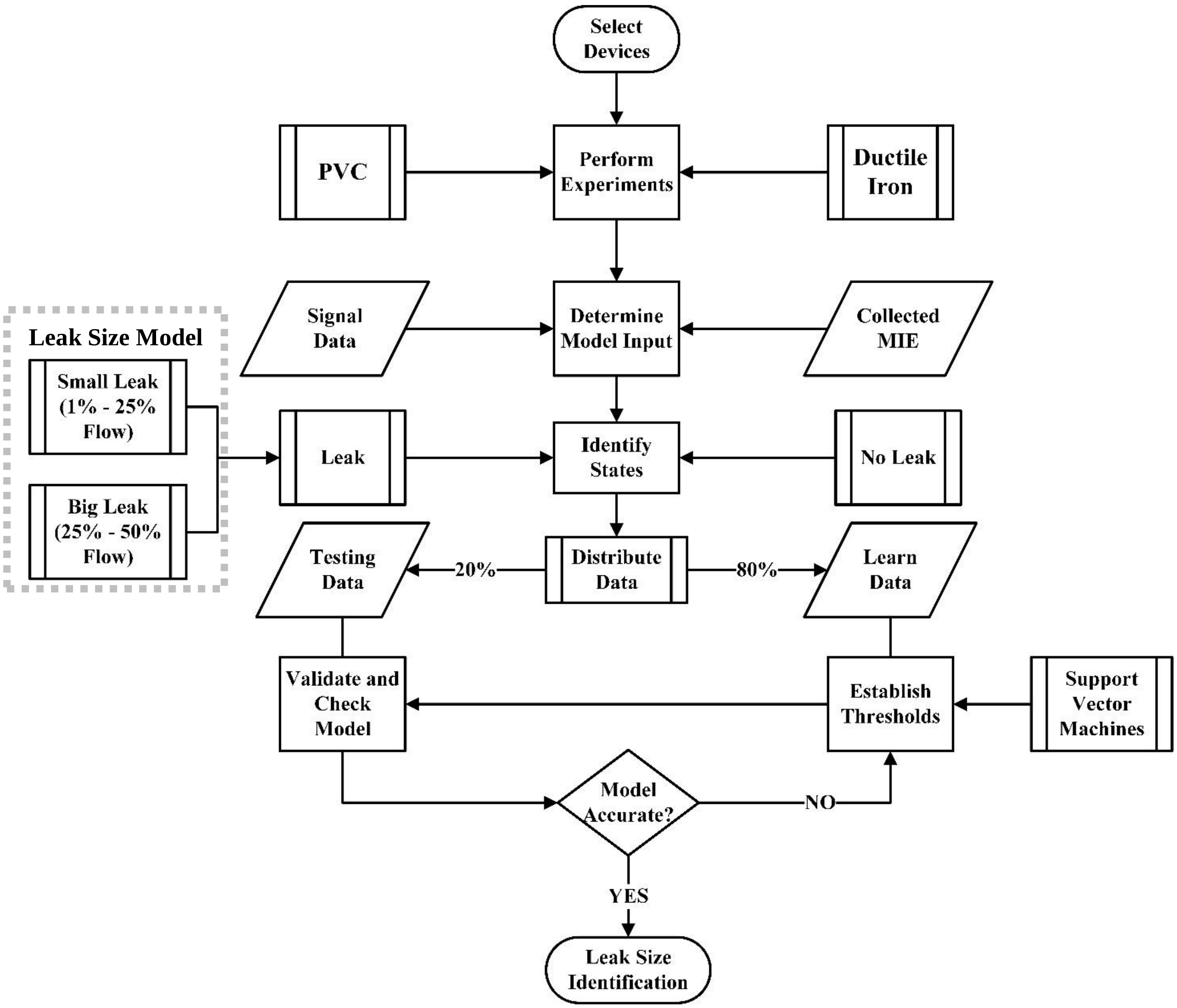

Figure 1: Leak Detection and Leak Size Classification Methodology

The states were identified and prepared for input. In order to develop this model, Linear SVM was utilized, whereas 80% of the data was used in model development and 20% were reserved for testing and validation. At the validation stage, the reserved data was utilized to determine the accuracy of the model. If the model was deemed accurate by means of having a very high correct predictions ratio, the model would be utilized. In the case of the model being unable to detect most of the cases properly, the input data and the model will be reassessed and the model will be redeveloped to eventually arrive at the required accuracy.

The development of a leak size prediction model is similar to the development of the leak identification model. As shown in Figure 1, the main differences lie in the way the data was inputted and the technique by which the experiment was performed. For the development of this model, instead of two states, there were three states, namely: No Leak, Small Leak, and Big Leak. With a small leak being between 1 to 25% of the total flow rate, and a big leak is estimated at 25% to 50% of the total flow rate. The dotted gray box in Figure 1 will be added to the original model in order to create a leak size classification model. SVM will be the development technique of this model as well.

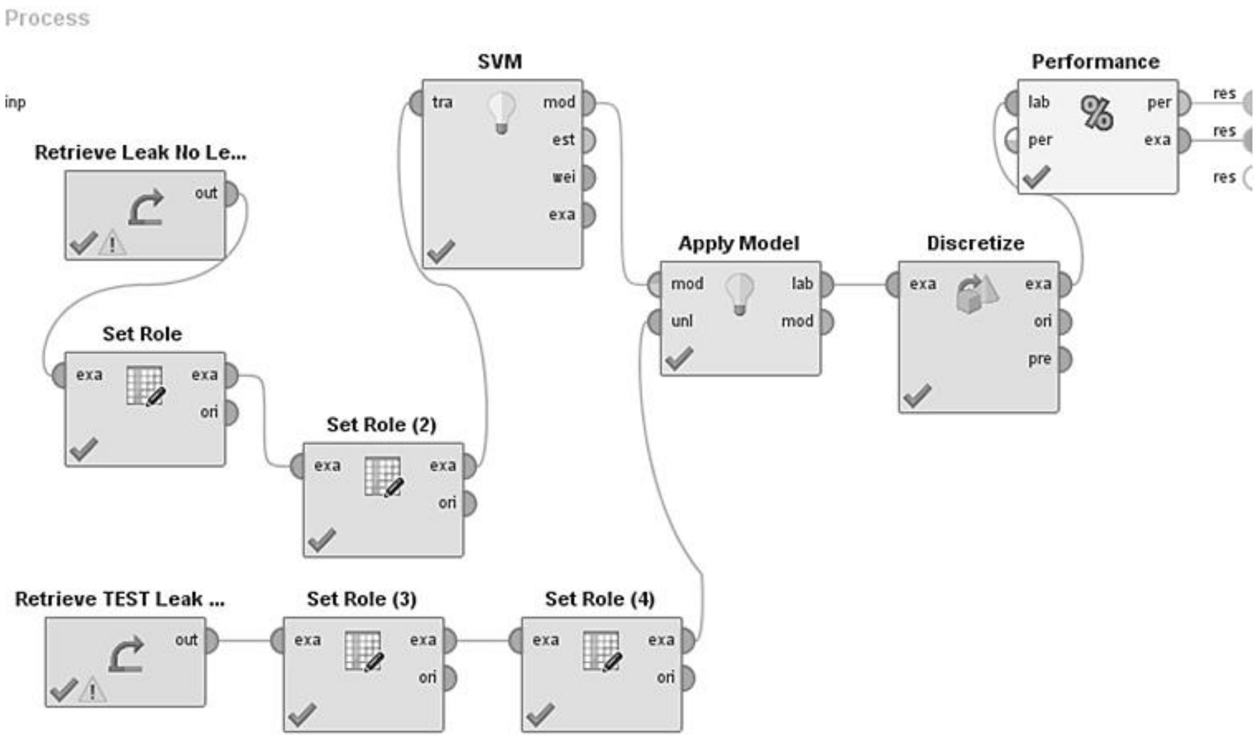

Figure 2: SVM Classification Model Architecture in RapidMiner 7.12

On the level of software data input and model architecture, Figure 2 visualizes the model design in RapidMiner 7.12. The data was first divided into two sets: (1) set 1 contained the training data and amounted to 225 data points of the format (MIE = 2.1, State = Leak), and (2) set 2 contained the validation data and was composed of 57 data points which are equal to 20% of the available data. After inputting the data, the required role of each input column was identified. The state was the main output value to be predicted, whereas MIE was the input for assessment. Training data was inputted into the main SVM function to develop the model and then both the developed model and the validation data were tested in the "Apply Model" function. The output of the model testing was discretized in order to test the performance of the model.

4. EXPERIMENTATION AND DATA COLLECTION

4.1 Experimentation

In order to understand the reaction of accelerometers in the presence of leaks, multiple experiments were required on multiple levels. The first level was carried out to understand the readings noted by the MEMS devices. Thus, at this stage, the first step was to setup any pipeline and place multiple sensors over the pipeline. **Error! Reference source not found.** displays the general setup of all experimental pipelines along with the relative distances and sensor placements. Each point noted as P(n), is a valve that can be opened at multiple values in order to simulate the leak. Pressurized water will be inserted through point P1 and will exit through point P7. The image in **Error! Reference source not found.** displays a two-inch ductile iron pipeline supported by two concrete blocks prior to the installation of the accelerometers and testing. The valves will act as leak simulators as they are slowly opened and closed.

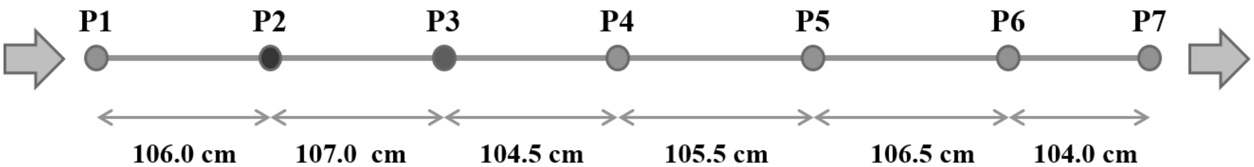

Figure 3: Experimentation Setup Diagram

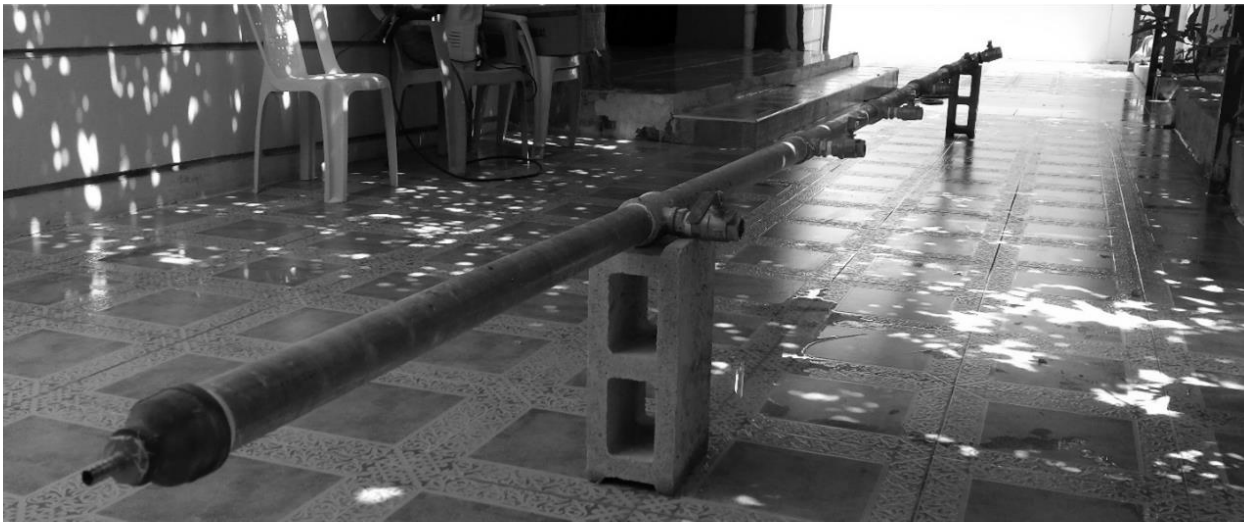

Figure 4: Two-inch Ductile Iron Pipeline

Throughout experimentation, it was noticed that the existence of an immediate open valve at the end can create violent signals that may propagate throughout the whole body of the pipeline and disrupt the data collection process. Thus, the solution was to create a damper that would allow the water to flow outside of the pipeline without creating a violent vibration that would disrupt experimentation. **Error! Reference source not found.** shows the solution in terms of a hose extension connected to the exit.

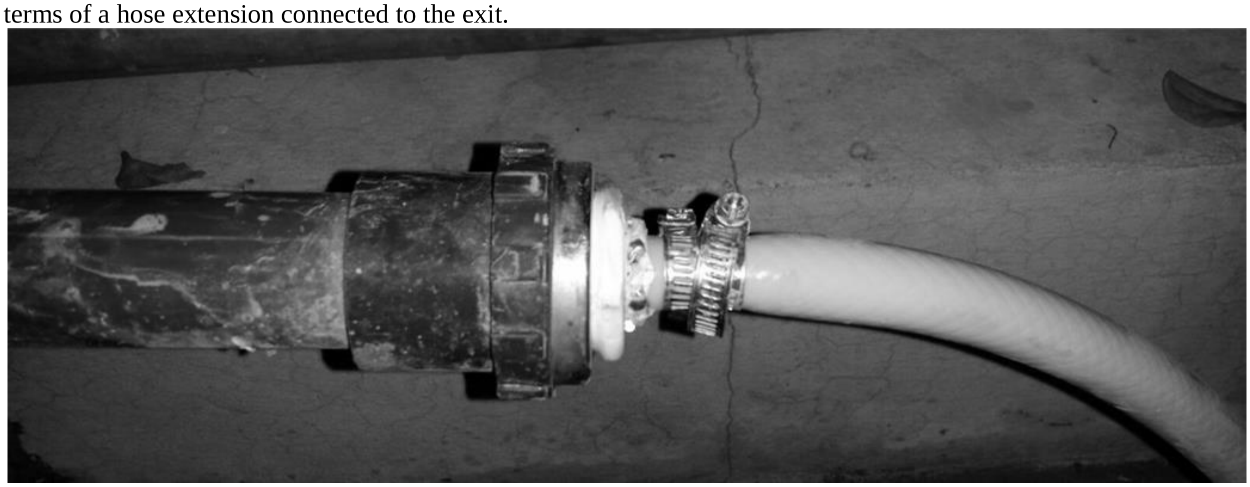

Figure 5: Pipeline Exit with Release Extension

After the general installation of the pipeline, **Error! Reference source not found.** displays how the sensors would be placed on each valve prior to the initiation of the water flow. Once water begins to flow within the pipeline, a certain amount of time is required for it to reach a pressurized state and maintain a uniform flow. Once the inlets and outlets have cohesive uniform flow, the experiments can be commenced. The figure also displays on the right side one of the performed experiments on a one-inch PVC pipeline, where a small leak is simulated via having a small opening in the valve.

Figure 6: One-inch and Two-inch PVC Pipelines

4.2 Data Collection

On the level of data collection, the accelerometers provided measurements of vibration signals in gravitational force units (g). Each second, these values were recorded for each accelerometer to monitor and identify the difference between states. **Error! Reference source not found.** displays the data sent by Sensor 1 during a one-hour on-off experiment. An on-off experiment allows the pipeline to behave normally with no leaks for the first five minutes and then a leak would be induced under the sensor for five minutes, then turned off again, and then on again, until the timer reaches one hour. The figure shows that the duration with no leak tends to be more stable and cohesive. On the contrary, when leaks are induced, the signal would become highly unstable and turbulent, in addition to being much bigger in terms of value.

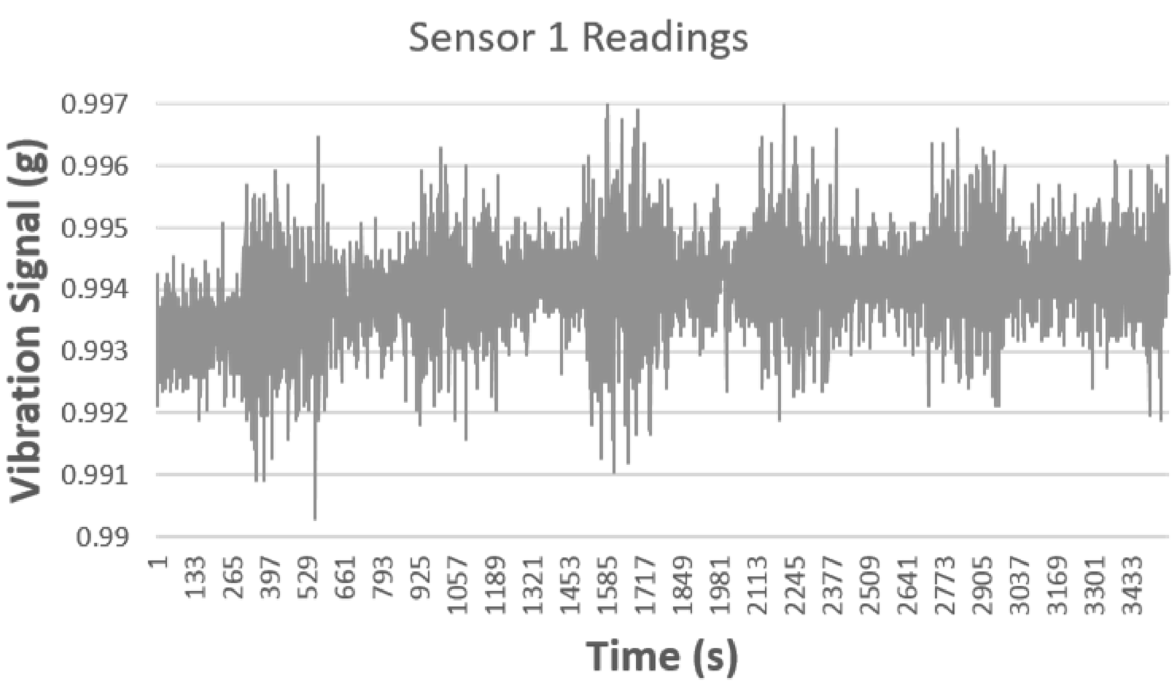

Figure 7: Sample of Received Signal during Experimentation

Using the previously mentioned model in **Error! Reference source not found.** that was developed by Martini et. al. in 2016, a Matlab code was developed. The code was designed to automate the process of data collection from sensors, compute all the necessary calculations to decipher the received signal data, and present them in terms of

MIE. The software takes in data sets of second by second readings and analyzes them to develop a bar chart that displays the condition of the pipeline by means of MIE for time span t=100 seconds. **Error! Reference source not found.** shows the result of the analysis performed to the signal received from sensor 1, as shown earlier in **Error! Reference source not found.**. The figure shows a clear distinction between the leak and no leak states within the signal, where the leak signal, displayed in black, has a much higher MIE value than that of the no leak signal, displayed in gray.

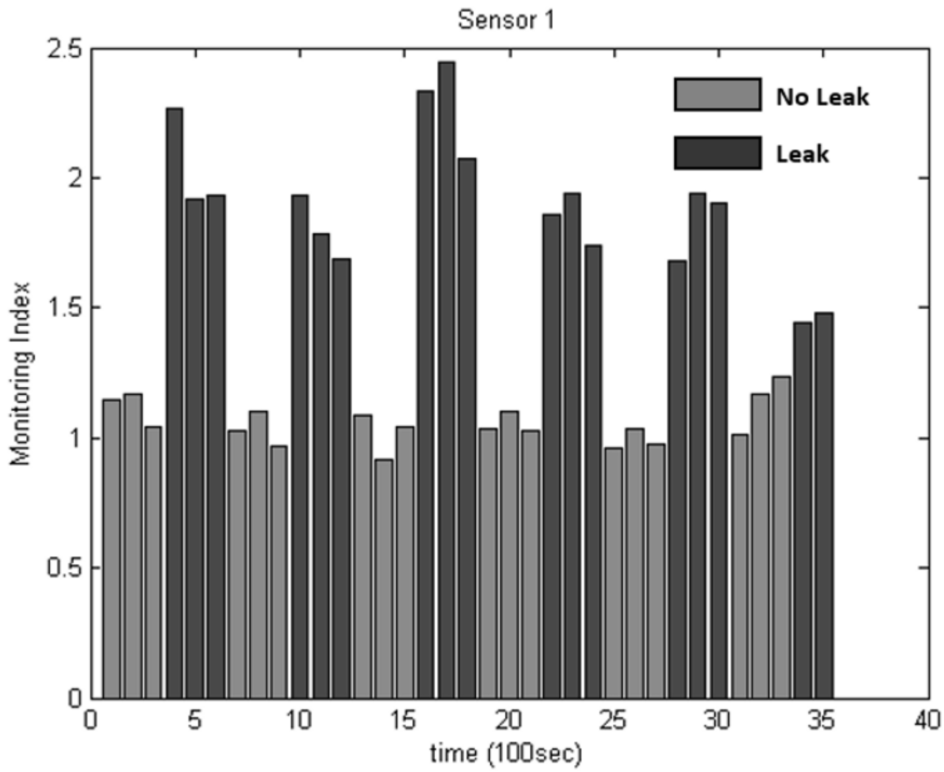

Figure 8: Sample Matlab Analysis of Sensor Signal

5. RESULTS

The first stage of the leak detection system revolves around developing a model that can classify the data and identify whether the pipeline is suffering from a leak or not. To develop this model, the linear SVM technique was utilized using 282 data points of the MIE in the leak and no leak states. The values were taken during experimentation from each sensor during the simulated leak and no leak states. After organizing the data into excel sheets and dividing the data into 80% learning data and 20% testing and validation data, the data was inputted to the Rapid Miner software version 7.12. Using the linear support vector machines function within the software, the data was classified based on their states and a threshold was identified. Table 1**: SVM Model Results** displays the output derived from Rapid Miner for the SVM model. The model achieved a 98.25% accuracy in terms of classifying and identifying leak states over 57 randomly selected testing points from the original data. The model had only one misclassification, which was a false alarm. Thus, the model was successfully able to identify the pipelines state. Furthermore, the model recommended an MIE threshold of 1.018 for leak detection. However, in order to eliminate the possibility of false alarms, it is recommended to increase that value to 1.15.

Table 1: SVM Model Results

Accuracy = 98.25%	Leak (True)	No Leak (True)	Class Precision

Leak (Predicted)	28	1	96.55%
No Leak (Predicted)	0	28	100%
Class Recall	100%	96.55%	

On the other hand, Table 2 displays the precision values of each class under study. The leak size classification model displayed an overall accuracy of 85%. Besides, the model was highly accurate in distinguishing no-leak states due to the 100% class precision as well as 100% class recall. Some loss of precision was identified while trying to distinguish between leak sizes, where big leaks had a 100% class recall, but a 60.87% class precision since the model predicted some small leaks as big leaks. Although small leaks had a class precision of 100%, the class's recall was at 40%, which shows that most of the confusion in the model was caused by the values emitted by small leaks. The size classification model has been more accurate in identifying the existence or non-existence of leaks than the leak identification model with 100% accuracy. Furthermore, the model has detected all big leaks very accurately, yet it confused most of the small leaks for big leaks. This confusion on the level of small leaks requires further investigation to identify the causes and facts behind it.

Table 2: Accuracy of Leak Size Classification Model

Accuracy = 85%	**Small Leak (True)**	**Big Leak (True)**	**No Leak (True)**	**Class Precision**
Small Leak (Predicted)	6	0	0	100%
Big Leak (Predicted)	9	14	0	60.87%
No Leak (Predicted)	0	0	31	100%
Class Recall	40%	100%	100%	

6. SUMMARY AND CONCLUSIONS

This paper presented a novel approach to identify the existence of leaks in pressurized pipelines using support vector machines along with micro-electro-mechanical sensors, specifically accelerometers. The presented model relies on measuring vibration signals in one-inch and two-inch during experimentation. The collected signal data was processed, analyzed, and accordingly fed to a support vector machine model. The results of the developed models displayed that the coupling of support vector machines with accelerometers showed promising results in providing accurate and significant input on the leaks' formation for PVC and ductile iron pipelines. Additionally, this coupling has the capability of detecting the size of leaks within a monitored pipeline.

The presented research can be further improved by identifying the rate of vibration signal decay within pipelines. Additionally, extra work should be directed towards determining the cause of the confusion in the small leaks identification. The model can be further enhanced through real-time application testing, utilizing other classification techniques, and using the model data for leak pinpointing.

7. REFERENCES

Alkasseh, J. M. A., Adlan, M. N., Abustan, I., Aziz, H. A., and Hanif, A. B. M. (2013). Applying minimum night flow to estimate water loss using statistical modeling: A case study in Kinta Valley, Malaysia. *Water resources management*, Springer, 27(5), 1439–1455.

Almeida, F. C. L., Joseph, M. J. B. P. F., Dray, S., Witfield, S., and Paschoalini, A. T. (2014). An investigation into the effects of resonances on the time delay estimate for leak detection in buried plastic water distribution pipes. *Eurodyn 2014: Ix International Conference On Structural Dynamics*, 3129–3136.

AWWA. (1987). *Leaks in water distribution systems: a technical/economic overview*. AWWA, Denver, Colorado, USA.

Cheong, L. C. (1991). Unaccounted for water and economics of leak detection. *Proc. 8th International Water Supply Congress and Exhibition*, International Water Supply Association, Copenhagen, 11–16.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine learning*, Springer, 20(3), 273–297.

El-Abbasy, M. S., Mosleh, F., Senouci, A., Zayed, T., and Al-Derham, H. (2016). Locating Leaks in Water Mains Using Noise Loggers. *Journal of Infrastructure Systems*, American Society of Civil Engineers, 4016012.

El-Zahab, S., Mosleh, F., and Zayed, T. (2016). An Accelerometer-Based Real-Time Monitoring and Leak Detection System for Pressurized Water Pipelines. *Pipelines 2016*, ASCE, Kansas, MO, 257–268.

Fletcher, T. (2009). Support vector machines explained. *Tutorial Paper*, <http://www.tristanfletcher.co.uk/SVM Explained.pdf> (Jan. 25, 2017).

Hamilton, S. (2009). ALC in Low Pressure Areas - It can be Done."*Proceedings of Water Loss 2009 South Africa*, Capetown, South Africa, 131–137.

Krchnak, K. (2016). Water Scarcity. *Water Scarcity - World Wild Life*, <https://thewaterproject.org/water-scarcity/> (Dec. 18, 2016).

Mandal, S. K., Chan, F. T. S., and Tiwari, M. K. (2012). Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM. *Expert Systems with Applications*, Elsevier Ltd, 39(3), 3071–3080.

Martini, A., Troncossi, M., and Rivola, A. (2015). Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements. *Shock and Vibration*, 2015, 1–13.

Mashford, J., De Silva, D., Marney, D., and Burn, S. (2009). An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. *Third International Conference on Network and System Security, 2009. NSS'09.*, 534–539.

Royal, A. C. D., Atkins, P. R., Brennan, M. J., Chapman, D. N., Chen, H., Cohn, A. G., Foo, K. Y., Goddard, K. F., Hayes, R., Hao, T., Lewin, P. L., Metje, N., Muggleton, J. M., Naji, A., Orlando, G., Pennock, S. R., Redfern, M. A., Saul, A. J., Swingler, S. G., Wang, P., and Rogers, C. D. F. (2011). Site Assessment of Multiple-Sensor Approaches for Buried Utility Detection. *International Journal of Geophysics*, Hindawi Publishing Corporation, 2011, 1–19.

Shinozuka, M., Chou, P. H., Kim, S., Kim, H. R., Yoon, E., Mustafa, H., Karmakar, D., and Pul, S. (2010). Nondestructive monitoring of a pipe network using a MEMS-based wireless network. *Proceedings of SPIE - Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*, 76490P.

Van Zyl, J. E., and Clayton, C. R. I. (2007). The effect of pressure on leakage in water distribution systems. *Proceedings of the Institution of Civil Engineers-Water Management*, 109–114.