Municipal Water Pipeline Leak Detection System

Hao Wang, Shiani Raj, Troy Lewis and Zhen Ye
Department of Chemical Engineering
University of Waterloo
Waterloo, Canada
h475wang@uwaterloo.ca, s5raj@uwaterloo.ca, t6lewis@uwaterloo.ca, zhen.ye@uwaterloo.ca

Haitian Zhang
PhD Candidate in the Department of Chemical Engineering
University of Waterloo
Waterloo, Canada
haitian.zhang@uwaterloo.ca

Hamid-Reza Kariminia
Department of Chemical Engineering
University of Waterloo
Waterloo, Canada
hrhamedaani@uwaterloo.ca

Ali Elkamel
Department of Chemical Engineering
Khalifa University, Abu Dhabi, UAE & University of Waterloo, Waterloo, Canada
aelkamel@uwaterloo.ca

Abstract

This work explores the feasibility of modernizing the current leak detection method used by the City of Kitchener which involves manual acoustic readings performed on a third of the city annually. We seek to develop software to detect leaks in real time using pressure and flowrate data collected by remote sensors in water pipelines. The primary objective is to update the detection to be in real time and increase sensitivity in the process by detecting smaller leaks that could have previously gone undetected. We have decided to achieve this using a time-series classification algorithm: MLSTM-FCN and the LeakDB dataset to represent a scaled-down version of the water distribution network in the City of Kitchener. The configuration of using pressure sensors only was selected from the results of the reduced feature test. It provided satisfactory performances in the proceeding generalization and localization tests. The solution fulfills all constraints and criteria. Based on the analysis, it is recommended to install 388 pressure sensors in the City of Kitchener as it minimizes the cost, without sacrificing the accuracy of the model.

Keywords
Leak detection, Time-series classification, MLSTM-FCN and Reduced feature test.

1. Introduction
Municipal water pipelines are important infrastructure in the City of Kitchener as they provide water to many residents, companies, and industrial plants in the city. Over time, water can leak out of these pipelines due to corrosion or old age, as the pipe material starts to degrade (Cody et al. 2020). Water leakages can often go undetected for long periods of time, leading to large water losses of up to 7 million liters of water in a day within Kitchener (WaterWorld 2023). The percentage of water lost due to leaks varies by region, with an average estimated to be up to 22% of total volume (Cody et al. 2020). The most common type of water leakage detection currently being used for municipal water pipelines is an acoustic method which is based on the sound of water within the

pipes (Khulief et al. 2012). However, this method is ineffective when detecting small leaks (Bakhtawar and Zayed 2021). The demand for high-accuracy water pipeline leak detections systems has been increasing in recent years due to the improvement in AI techniques.

2. Problem Statement and Specifications

2.1 Problem Statement

The City of Kitchener currently relies on a traditional acoustic method for leak detection and localization, which is costly, requires heavy labor, and is not real-time: only 1/3 of the city is tested each year. Many leakages cannot be detected until they surface and not only cause a loss of up to 30% of water supply but also affect residents and companies as their water supply may be cut for a prolonged period until the leakage is fixed, interrupting daily life and operations.

2.2 Criteria and Constraints

The three identified criteria that must be met are ease of operation, system reliability, and project costs, which are shown in Table 1. After the project is implemented, specially trained operators will be required for the daily operation of the software. A detailed training manual will need to be developed to train operators on the specifications of this software. Most operators will only be required to know front end details of the software and training can be expected to only take a few days. It would be advised to have one lead operator trained to troubleshoot the software if problems arise. This role would not be considered normal operation and the complexity of it is not included in the ease of operation criteria. Regarding the reliability of the model, it is not enough to simply detect leaks, but it must correctly identify leaks at a high degree of accuracy, which is set to be 90%. This includes not only correct detection of leaks but also not falsely flagging leaks when there is not one present. The minimum accuracy level will be achieved by a variety of sensor setups. This allows for flexibility in determining the best sensor setup, which will be explained in further detail later in this report.

Table 1. Description of criteria

Criteria	Description/Explanation
Ease of Operation	Training of employees on new systems should not be lengthy or strenuous
System Reliability	The new system should consistently and correctly detect leaks with an accuracy > 90%
Project Costs	The implementation and operation costs of the new system should be minimized

Table 2. Description of constraints

Constraints	Description/Explanation
Localization Ability	The municipal water pipeline leak detection system must locate leaks within 5 meters of the actual leak point
Generalization Ability	Percent difference in fault detection rate at each potential leak node should be less than 20%
Maximum Operating Pressure (MOP)	The system must monitor and alert the user when the municipal water pipeline network is being operated near or above MOP (100 psi)

As shown in Table 2, the three constraints that should be met are the localization ability of the leak detector model, the generalization ability, and the final constraint being the maximum operation pressure of the network. For localization, nodes are established in the theoretical network, and leaks should be identified by the nearest node. Upon detection at a node, traditional methods should be utilized to confirm the presence of a leak and narrow down its location. The pipeline network in Kitchener has pipe diameters ranging from 4” to 30” and so our project needs to properly operate anywhere in the distribution system regardless of the pipe diameter. To ensure generalization is met, a range of diameters are used in the simulated network sample to verify acceptable detection rates. The final constraint is the maximum operating pressure. There are already pressure sensors installed at many locations throughout the distribution system, however they do not transmit the data in real time. As such, upgrading these sensors and installing additional sensors where needed must not impact the operating pressure of the pipeline network.

3. Solution Selection

3.1 Potential Solutions

- Use Multivariate LSTM-FCN (MLSTM-FCN), a time-series classification neural network in Deep Learning, to build a real-time model-free leak detection system. Leak detection is achieved by detecting the leak patterns in the pressure and flowrate data (Punukollu et al. 2022).
- Use You Only Look Once (YOLO) Version 5, a real-time object detection neural network in Deep Learning, to build a real-time model-free leak detection system. Leak detection is achieved by detecting leak patterns in sound waves (acoustic signals) (Cody et al. 2020).
- Increase inspection and maintenance frequency of current leak detection method to prevent leakage from happening.

3.2 Weighted Decision Matrix

Reference Solution: The existing method used in the city is the traditional acoustic leak detector, which covers about 1/3 of the city per year. The corresponding criteria solution screening is illustrated in Table 3.

Table 3. Criteria solution screening

Constraint/Criteria List	Weight (%)	Potential Solutions							
		Ref		1		2		3	
		R	S	R	S	R	S	R	S
Localization ability	20	0	0	-1	-20	-1	-20	0	0
Generalization Ability	20	0	0	2	40	0	0	0	0
Maximum Operating Pressure (MOP)	20	0	0	1	20	1	20	0	0
Ease of Operation	10	0	0	-1	-10	-1	-10	0	0
System Reliability	15	0	0	2	30	1	15	1	15
System Cost	15	0	0	1	15	1	15	-2	-30
Total	100	0		75		20		-15	
Rank				1		2		3	
Selected				Yes		No		No	

* Note:

- Localization ability: in Solutions 1 & 2 the localization precision is limited by the sensor distribution within the network, and thus poorer performance than traditional acoustic leak detector (Solution 3 and reference solution).
- Generalization ability and system reliability: reference solution and Solutions 2 & 3 are acoustic methods that are subjected to background noise (Cody et al. 2020), but Solution 1 isn't affected by background noise. Thus, reference solution and Solutions 2 & 3 are less reliable (system reliability) and can't be applied to all sections of the pipeline network (generalization ability) compared to Solution 1.
- Maximum Operating Pressure: reference solution and Solution 3 require manual leak detection while Solutions 1 and 2 are real-time detection. So, Solutions 1 & 2 are better than Solution 3 and reference solution.
- Ease of operation: Compared to reference solution, Solutions 1 & 2 require employees to be familiar with the new system, but Solution 3 doesn't. Thus, Solution 3 is better than Solutions 1 & 2.
- System cost: Solutions 1 & 2 cost less compared to reference solutions, since algorithms are open-sourced and able to use the existing sensors for measurement. Solution 3 increases inspection and maintenance frequency, which will increase the cost dramatically compared to reference solution.

4. Technical Solution Analysis

As shown in Figure 1 below, the solution analysis consists of three main parts: data pre-processing, training, and testing.

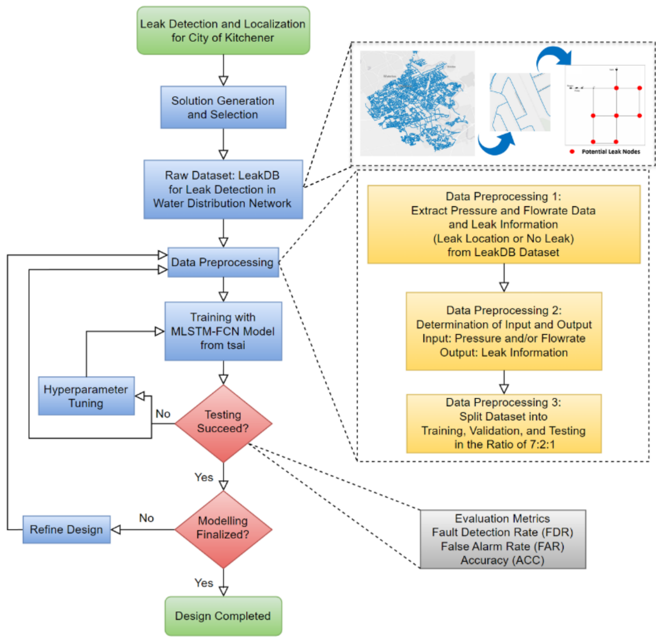


Figure 1. Design process flowsheet

4.1 Data Pre-Processing

The City of Kitchener possesses a complex water distribution network, as illustrated in Figure 1 above. Unfortunately, real data from the city was not obtained. As a result, the LeakDB dataset was selected to represent a scaled-down version of the water distribution network in the City of Kitchener. It is a benchmark dataset for leak detection in water distribution networks with 1000 different artificial scenarios. Each scenario contains pressure data, flowrate data and leak information such as whether the network is with leaks or without a leak (Vrachimis et al. 2018).

Before the data pre-processing of the LeakDB dataset, the following assumptions and approximations are utilized:

- Assume only one leak can simultaneously occur in the water distribution network.
- If a leak occurs, assume the leak continuous for at least 48 hours (Kammoun et al. 2022).

With the assumptions listed above, three steps were taken for data pre-processing. First, pressure and flowrate data as well as leak information were extracted from the raw LeakDB dataset. Second, use the extracted pressure and flowrate data as an input and leak information as the corresponding output to construct a customized dataset. The leak information was presented as different classes for classifications. A total of eight classes were used: leak at Nodes 12, 13, 21, 22, 23, 31, 32 and no leak. After filtering out the scenarios that don't satisfy the two assumptions listed above, a total number of 493 scenarios were selected to build the customized dataset. Each class contains 50 to 73 samples, leading to a relatively small and imbalanced dataset. The final step divides the customized dataset into training, validation, and testing sets in a ratio of 7:2:1.

4.2 Training and Testing

For training and testing, the assumptions adopted are listed below.

- Assume a "pressure at a node" or "flow at a link" represents a "sensor" measurement, which is also referred to as a "feature" in this project.
- Approximated leak location is at each potential leak node (red dots in Figure 1)

However, these two assumptions lead to the problem that the leak detector's localization precision is limited by sensor distribution in the water distribution network. Further statistical analysis methods can be applied to increase localization precision. For example, the variation in measurement at each sensor location is related to the relative distance to the leak point. The closer to the leak point, the greater the variation. Thus, the leak location can be estimated by statistical distribution analysis regarding the variation in sensors' measurements and their corresponding locations (Kammoun et al. 2023).

4.2.1 Time-series Classification

In the solution selection section, time-series classification using the MLSTM-FCN algorithm was identified as the selected solution. As illustrated in Figure 2 below, time-series classification is an engineering tool that learns patterns and their associated classes like "leak at node 3" in Graph C or "no leak" in Graph B from data (training) and uses this knowledge to predict the class associated with new data in Graph A (testing or leak detection).

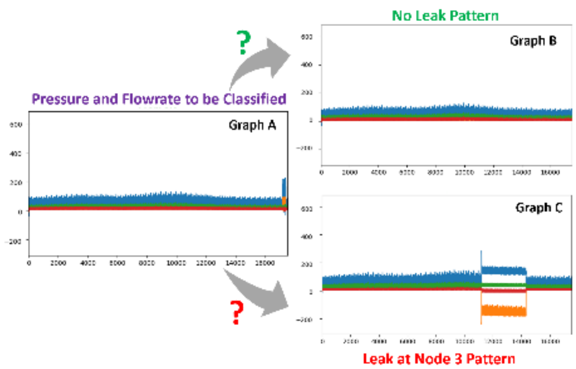
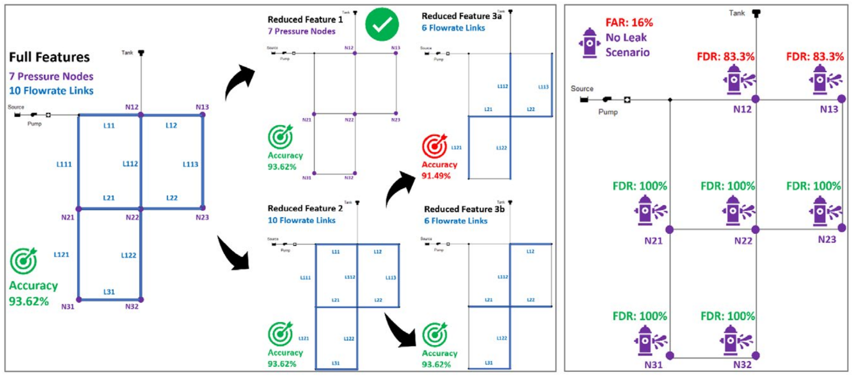


Figure 1. Working mechanism of time-series classification

The training process utilized the training and validation sets, as well as the MLSTM-FCN algorithm from tsai, a state-of-the-art Deep Learning library for Time Series and Sequences. MLSTM-FCN is a widely adopted multivariate time-series classification algorithm in Deep Learning in the field of Artificial Intelligence. It extends the popular univariate time-series classification algorithm: LSTM-FCN to multivariate with the addition of a squeeze-and-excitation block to improve accuracy. Multivariate means multiple variables or sensors' measurements. This means it is well-suited for leak detection tasks with multivariate inputs like pressure and flowrate data. Moreover, the MLSTM-FCN model provided by tsai is open-sourced (free), which significantly reduces the project cost.

4.2.2 Reduced Feature Test

Due to the assumption that a "feature" represents a "sensor", the number of features used for leak detector training and testing is equal to the number of sensors that will be placed in the water distribution network. The lower the number of features utilized, the lower the number of sensors that are required, and thus the lower the project cost. Hence, a reduced feature test was conducted to explore the potential and feasibility of sensor reduction.
The test results were obtained and illustrated in Figure 3a below. As observed from Figure 3a, the system reliability criteria are satisfied in all cases considered as they all resulted in an accuracy greater than 90%. For feature reduction, reducing the number of features (sensors) to pressure data only (Reduced Feature 1) or flowrate data only (Reduced Feature 2) is feasible without losing accuracy achieved at full feature: 93.62%. However, further decreasing the number of flowrate data used requires case-by-case analysis since it leads to inconsistent accuracy. This is due to that when using 6 flowrate links, Reduced Feature 3b led to a decreased accuracy of 91.49% while the accuracy for Reduced Feature 3b remained at 93.62%. Thus, Reduced Feature 1: 7-pressure node, was selected as the configuration adopted for further analysis of the leak detector's generalization and localization ability.



(a) (b)

Figure 2. Test results for (a) reduced feature test and (b) generalization and localization ability test on Reduced Feature 1

4.2.3 Generalization and Localization Ability Test

The generalization and localization ability tests were conducted under Reduced Feature 1: using pressure data from seven pressure nodes to train and test the leak detector. Two evaluation metrics are of interest for this test: (1) fault detection rate (FDR) for each potential leak node and (2) false alarm rate (FAR) for no-leak scenarios (Swain et al. 2020). They are given as follows:

$$FAR = \frac{\#\ of\ No\ Leak\ Scenarios\ Wrongly\ Detected\ as\ Leak}{\#\ of\ No\ Leak\ Scenarios}$$

$$FDR = \frac{\#\ of\ Correctly\ Detected\ Leaks\ at\ Node\ i}{\#\ of\ Leaks\ at\ Node\ i}$$

The test results were displayed in Figure 3b above. As illustrated in Figure 3b, leaks are detected and localized at each potential leak node, which limited the localization precision. It can be improved using the statistical analysis methods mentioned at the beginning of this section. Although the percent difference in fault detection rate between leak nodes is 16.7%, less than the 20% generalization constraint. The leak detector isn't well-performed since the FDR at Nodes 12 and 13 are only 83.3% while the FAR for no leak scenarios is 16%. This might be because of that the customized dataset used is small and imbalanced. Time-series data augmentation methods such as reversion had been utilized to increase and balance the number of samples in the dataset. However, this altered the sample distribution within the dataset (data bias) (Xu et al. 2020) and resulted in unreasonable trends observed in the reduced feature test.

5. Sustainability (Impacts) Analysis

5.1 Environmental Impact

30% of energy used by The Region of Waterloo, the source for Kitchener's treated water, is used in the treating and pumping of water (Region of Waterloo 2019). In the year 2018 the electricity used in the treatment of water was 26,295 MWh, and for pumping and distribution was 14,143 MWh. Given an emission factor of 30 gCO_2e per kWh, this results in 1,213 tons CO_2e produced annually in water treatment and distribution throughout the region. The regions consist of three cities, Kitchener, Waterloo, and Cambridge, plus several small towns such as Elmira, Wellesley, and Woolwich. Approximately one hundred million liters of water, or 100,000 m^3, is pumped throughout The Region daily, and anywhere from 12% to 40% of this water is lost depending on the area (Region of Waterloo 2019). A conservative estimate of the percentage of water used by the city of Kitchener would be 25%. This puts the daily amount of water pumped throughout the city of Kitchener at roughly 25,000 m^3. The city of Kitchener is divided into seven sections where water flow is measured, taking an average percentage lost for each of these sections a value of 28% lost water is estimated. This brings the volume of water lost throughout the city of Kitchener to approximately 7,000 m^3 per day. This volume accounts for 7% of the daily volume of water pumped throughout The Region. If an assumption is made that 7% of the energy consumption goes towards the lost water, then 2,831 MWh would be wasted. However, as any water reduction would be a reduction of water drawn from The Grand the treatment requirement would be higher than average. The treatment process can be broken down into three components. First, transportation from the source to the treatment station, which requires 0.002 kWh/m^3. The main treatment process, requiring 0.75 kWh/m3. Lastly, pumping the treated water into the distribution network, which requires 0.68 kWh/m^3 (Plappally and Lienhard V 2012). With the previously discussed lost water of 7000 m^3

then for transportation, treatment, and pumping, the energy squandered is 1916 MWh, 26 MWh, and 1737 MWh respectively. At an emission factor of 30 g CO_2e/kWh this accounts for 110 tons CO_2e (Government of Canada 2019). Identification and repair of leaks in the pipeline network will take time and as such these losses will not be recognized as reduced in full for many years. If the first-year reduction is taken to be 50% then the emission reduction would be 55,000 kg CO_2e, with a steady increase to 90% reduction as seen in Figure 4 below.

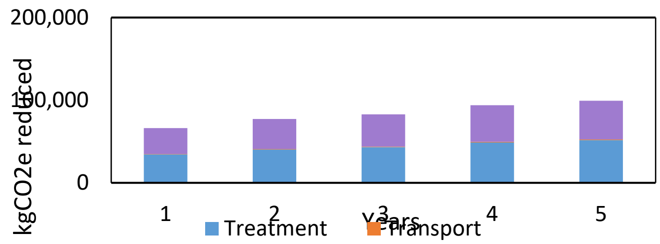


Figure 3. Annual GHG emission reduction

The emissions that would be generated in comparison to the reduction are extremely small and could justifiably be considered negligible. The main new source of emissions would be the operation of the hardware required to run the model. To ensure enough processing power it would be recommended that each of the seven mentioned sections of the city have a dedicated processing server, which would require a 1500W power supply. It is within reason to assume an 80% efficiency rating on the power supply, so then an annual power requirement of 16,425 kWh annually. This would be 493 kg CO_2e annually, which is less than 1% of the projected first year emission reduction. Once leaks are being identified and repairs are being performed there will be an increase of emissions associated with the repair process. That is, however, not a result of this project. The leaks are not a result of this project, only identified by the project. The leaks being repaired would need to be repaired eventually, this project is only expediting the process. As such, any emissions associated with the repair of leaks are not caused by a successful implementation.

5.2 Economic Impact

The City of Kitchener has a water distribution network with pipeline infrastructure totaling approximately 890 km (SCG Flowmetrix 2021). Our model uses a scaled-down water distribution network that only analyzes 16 km of pipeline. For the scaled-down version, as mentioned in the solution analysis, 4 different feature options were investigated, and Table 4 demonstrates the reduced feature accuracy for each type.

To determine the best model, the criteria outlining project costs must be satisfied such that the total cost is minimized without sacrificing accuracy. The total savings from reduction in lost water and energy from the treatment process must be greater than the implementation cost. The 4 model types will be compared based on the implementation cost and payback period, which is the time it takes to recover the initial investment.

Table 4. Reduced feature accuracy

Type	Number of Pressure Sensors	Number of Flow Meters	Associated Accuracy
Type 1: Full Features	7	10	93.62%
Type 2: Reduced Features – Pressure Only	7	0	93.62%
Type 3a: Reduced Features – Flowrate Only	0	10	93.62%
Type 3b: Reduced Features – Flowrate Only	0	6	92.60%

The implementation cost includes the purchasing cost of all the pressure sensors or flowmeters for the City of Kitchener, the new hardware equipment needed to run the detector model, and the installation costs associated with newly purchased sensors, which are illustrated in Table 5.

Both the flowmeter and pressure sensors need to precisely record the flowrate and pressure of water in underground pipelines and send signals wirelessly to the main control system for data collection. The collected data will then subsequently be used to run the detector in real time. The flowmeter and pressure sensor models both use HART technology so that there is wireless communication of data between the sensors and control systems. The cost of the required flowmeter and pressure sensor were obtained through quotes priced to be $950 (ITM Instruments INC 2023) and $837.50 (VERIS 2023) per sensor, respectively. Since the number of sensors are known for the 16 km of pipeline analyzed in our model, an extrapolation method was used to determine how many sensors were required for 890 km of pipeline to calculate the total sensor costs. Based on the number of sensors required for each feature type, the labor costs associated with installing all these sensors will be calculated as well using hourly wages of City of Kitchener 's water utility workers (Job Bank 2023).

Table 5. Reduced feature costs

Type	Pressure Sensors	Flow Meters	Implementation Cost
Type 1: Full Features	389	556	$ 2,300,000
Type 2: Reduced Features – Pressure Only	388	0	$ 973,000
Type 3a: Reduced Features – Flowrate Only	0	556	$ 1,400,000
Type 3b: Reduced Features – Flowrate Only	0	334	$ 880,000

Our model currently only requires the use of one GPU server. A GPU server is a computational server that is used for data processing and the system which the detector-model runs on. To be able to have enough data processing power for 890 km of pipeline, 10 servers will need to be purchased by the city and this cost will also be accounted for in the implementation cost (GIGABYTE 2023). More detailed sample calculations are shown in the Appendix.

Based on the implementation cost, Type 3b has the lowest capital cost, due to its low number of flowrates, closely followed by Type 2 which only uses pressure sensors in its model. While the design must minimize capital costs, it also must be lower than the total savings from preventing water loss.

From Section 5.1, it is known that approximately 7000 m^3 of water is lost in the City of Kitchener per day. The consumer water rate in the City of Kitchener is $2.62/$m^3$ (Utility rates - City of Kitchener 2023) and so the total cost of the water loss to the City of Kitchener is $18350 per day. In addition, from Section 5.1, it was determined that the total energy wasted due to treatment, transport and pumping of water was 10080 kWh per day. Using the city's energy rate of $0.07/kWh, the city's annual cost of water loss is $7 million.

By implementing our leak detector model, with accuracy of the Type 2 and Type 3b leak detectors being greater than 90%, there is potential savings of $6.3 million per year, without accounting for the operating costs of the leak detector. After year 0 in which all the equipment is purchased and installed, from year 1 onwards, there will be annual operating costs which include labor costs for running the actual leak-detector software and annual costs for maintenance which involves city personnel checking the sensors and ensuring they are well calibrated. Using the same extrapolation method as previously, for 890 km of pipeline, 10 workers are needed to monitor and operate the leak-detector model. Using the hourly wages of City of Kitchener's water utility workers (Job Bank 2023), the annual operating costs are determined to be $880,000.

The payback period was calculated based on an analysis of 5 years. The cumulative cash flow for each year was determined by taking the difference of potential savings and annual operating costs. By doing so, the break-even point of the investment was ascertained as shown in Table 6.

Table 6: Reduced feature payback period

Type	Capital Cost	Payback Period	Associated Accuracy
Type 1: Full Features	$ 2,300,000	2.3 years	93.62%
Type 2: Reduced Features – Pressure Only	$ 973,000	1.3 years	93.62%
Type 3a: Reduced Features – Flowrate Only	$ 1,400,000	1.7 years	93.62%
Type 3b: Reduced Features – Flowrate Only	$ 880,000	1.4 years	92.60%

The lowest payback periods are for Type 2 and Type 3b models. While the implementation cost of Type 3b is a bit lower than the Type 2 implementation cost, it is important to note that using only pressure sensors will provide better accuracy. As such, installing 388 pressure sensors will lower the cost of this project and increase the return on investment.

5.3 Safety

With the implementation of our software, the city of Kitchener can achieve real time detection of pipeline leakages. In this way leakages can be detected before they can surface or cause sinkholes and prevent these safety issues from escalation. However, implementing the detection software also means that future detection will be heavily relying on the software. If the detection model failed or experienced any technical difficulties, such as power outage or service disconnection, water leakages may be undetected for a longer period of time. If the model fails without notice, leakage will not be detected until it surfaces. This could ultimately result in worse incidents than with current methods.

Some potential safety issues may also occur during the implementation stage of the leakage detection system. Currently, pressure sensors in the pipelines of City of Kitchener do have wireless transmission of data, which is needed for our leakage detection software. As such, updated pressure sensors will be installed in the water pipes across the city of Kitchener as part of the implementation process. As most of the pipelines are located on the roadside and underground, installing these instruments will mean roadworks, which poses risks of workplace injury and safety issues from traffic congestion.

5.4 Social/Cultural Impact

Implementation of our project would result in the automation of the leak detection process. This would see a reduction in the hours of the technicians performing the testing and the project manager that prepares the report. While this annual contract would be lost the city may wish to hire said company to investigate detected leaks. This project would also create new job opportunities, both in the initial implementation and in the continued operation. Installation of 388 sensors would require 3000 hours of labor. Maintenance on the new sensors would also need to be done regularly to ensure proper calibration. Operators will also be required to use the AI that was developed which will create up to seven full-time jobs.

Water scarcity is not something most Canadians are concerned about, but it is a growing concern for a large portion of the world. It is believed that India and China, as well as some countries in Europe and Africa, will face water scarcity by 2025 (Water Scarcity 2023). Kitchener draws roughly 20% of its drinking water from the Grand River, a surface water source, which is particularly vulnerable to climate change. An example of this vulnerability can be seen as recently as July with Lake Mead water levels becoming dangerously low due to unprecedented drought (Earth Observatory 2023). As previously mentioned, any reduction in water lost would be a direct reduction in water that would need to be drawn from The Grand, which would provide protection against potential climate change concerns in the future. As well, by reducing the water lost in the distribution system now it allows for continued growth of the city, both residentially and commercially.

6. Conclusions and Recommendations

From technical analysis, the solution selected for building the leak detector is time-series classification on pressure and/or flowrate data using the MLSTM-FCN algorithm in Artificial Intelligence. The LeakDB dataset was utilized. A reduced feature test was conducted to explore the potential and feasibility of reducing the number of sensors used. It was found that using pressure or flowrate data is only feasible without losing accuracy. However, decreasing the number of flowrate data used requires case-by-case analysis since it leads to inconsistent accuracy. All cases considered in the test resulted in an accuracy greater than 90%, satisfying the system reliability criteria. Using pressure sensors only was selected as the desired configuration and further analysis was performed to test its localization and generalization ability. The test results showed that the leaks were localized and detected at each potential leak node. The 20% generalization constraint was met with a 16.7% difference in fault detection rate between leak nodes. Nevertheless, the performance was limited by the relatively small and imbalanced dataset. Moreover, the current solution can achieve the 5-meter localization precision constraint, but it is significantly confined by the sensor distribution in the water distribution network. Additional statistical analysis methods will be required to further increase leak localization precision.

Overall, it was concluded that using a reduced features method of only pressure sensors will be the optimal solution. For the City of Kitchener, 388 pressure sensors will need to be installed. The return on investment is only estimated to be 1.5 years. Further technical analysis confirms that using pressure sensors only is feasible without sacrificing accuracy of the design.

For the next steps, we recommend performing analysis on a larger dataset that is more complex to scale up the representation of the water distribution network for the City of Kitchener. Since the model only analyzes 16 km of pipeline, a model that uses a dataset of more than 100 km will be able to better reflect how the pipelines in City of Kitchener behave. In addition, further developing the model to detect 2 leaks that occur simultaneously in the water distribution network will be more representative of how leakages occur in real-world situations.

Currently the pipeline network within the City of Kitchener is divided into seven subsections, where flowrate into each subsection is monitored and compared to billed water. However, these subsections are still very large and are not uniform in size, with the largest subsection representing over 200 kilometers of pipeline. At the time of collection, one specific area was losing 40% of water pumped into the area, but due to the size, it was not feasible to investigate ahead of the routine annual testing. By implementing our model and installing the proposed sensors, a more consistent view of the pipeline network will be achieved.

7. Appendix

7.1 Environmental Analysis Calculations

Using the estimated lost water of 7,000 m^3 an approximation of energy consumption that could potentially be saved is calculated. The treatment process can be broken down into three components, transfer to the facility, treatment, and distribution. Each of these phases has an associated power requirements as follows (Plappally and Lienhard V 2012):

Transport: $0.002\ \frac{kWh}{m^3 * km}$

Treatment: $0.75\ \frac{kWh}{m^3}$

Distribution: $0.68\ \frac{kWh}{m^3}$

Average distance for treatment facility from the Grand River is taken to be 5 kilometers, the daily energy used to treat lost water can be calculated to be:

$$Transport = 0.002\ \frac{kWh}{m^3 * km} * 5\ km * 7000 \frac{m^3}{day} * 365\ \frac{days}{year}$$

$$Transport = 25550\ \frac{kWh}{yr}$$

$$Transport = 25.6\ \frac{MWh}{yr}$$

$$Treatment = 0.75\ \frac{kWh}{m^3} * 7000 \frac{m^3}{day} * 365\ \frac{days}{year}$$

$$Treatment = 1916250\ \frac{kWh}{yr}$$

$$Treatment = 1916\ \frac{MWh}{yr}$$

$$Distribution = 0.68\ \frac{kWh}{m^3} * 7000 \frac{m^3}{day} * 365\ \frac{days}{year}$$

$$Distribution = 1737400\ \frac{kWh}{yr}$$
$$Distribution = 1737\ \frac{MWh}{yr}$$
$$Total\ Energy = 3679\ \frac{MWh}{yr}$$

With an emission factor in Ontario of 30 $kgCO_2e$ per megawatt-hour (Government of Canada 2023), the annual emissions from treating lost water can be calculated:

$$Annual\ Emissions = 3679\ MWh * 30\ \frac{kgCO_2e}{MWh}$$
$$Annual\ Emissions = 110370\ kgCO_2e$$

7.2 Economic Analysis Calculations

Shown below are sample calculations of the process taken to determine all the costs required to calculate the payback period. Note that only calculations for Type 2, which is pressure sensors only will be shown.

- Implementation Cost:

The price of each pressure sensor is \$837.50 (VERIS 2023), and an extrapolation can be done to determine the number of pressure sensors needed and the cost of purchasing sensors for 890 km if 16 km of pipeline needs 7 pressure sensors:

$$Number\ of\ Sensors = \frac{890\ km}{16\ km} \times 7\ sensors = 388\ sensors$$
$$Sensor\ Costs = 388\ sensors \times \$837.50 = \$326101.56$$

Based on estimates from the company Microsensor, it costs \$1050 (VERIS 2023) to install a single pressure sensor and requires 3 personnel to assist in the installation which takes up to 4.5 hours. As such, the installation cost based on the City's labor wage is (Job Bank 2023):

$$Installation\ Costs = 388\ sensors \times \$837.50 \times 3\ workers \times \frac{\$36.25}{hr} \times 4.5\ hr \times \$1050 = \$597277.5$$

As mentioned in section 5.2, the city will need 10 GPU servers that cost \$5000 (GIGABYTE 2023) and so the total hardware equipment cost is:

$$Hardware\ Equipment\ Cost = 10\ servers \times \$5000 = \$50000$$

Therefore, the total implementation cost is:

$$Implementation\ Cost = \$326101.56 + \$597277.5 + \$50000 = \$973379.06$$

- Operating Cost:

As mentioned, 10 workers are needed to run the detector throughout the year and so based on the city's hourly wages for utility workers (Job Bank 2023), the annual running cost of the detector is:

$$Detector\ Running\ Cost = 10\ workers \times \frac{\$41}{hr} \times 8\frac{hr}{day} \times 5\frac{day}{week} \times 50\frac{weeks}{year} = \$820{,}000$$

In addition, the sensors must be maintained and re-calibrated every year so the maintenance costs per year is:

$$Maintenance\ Cost = 388\ sensors \times 3\ workers \times \frac{\$36.25}{hr} \times 1\ hr = \$42195$$

Therefore, the total operating cost is:

$$Operating\ Cost = \$42195 + \$820{,}000 = \$862195$$

- Savings:

As mentioned in section 5.1 and 5.2, by implementing our model, the potential savings is the cost associated with the City of Kitchener's water loss:

$$Savings = \left(\frac{\$2.62}{m^3} \times 7000\ m^3 + 10080\ kWh \times \frac{\$0.07}{kWh}\right) \times 365\ days = \$6970000$$

- Payback Period:

Table 7: Payback period

Year	Net Cash Inflow	Cumulative Cash Flow
0	$-\$973379.06$	$-\$973379.06$
1	$(\$6970000 \times 0.6) - \$862195 = \$3319805$	$-\$973379.06 + \$3319805 = \$2345622$
2	$(\$6970000 \times 0.7) - \$862195 = \$4015867$	$\$2345622 + \$4015867 = \$6361489$
3	$(\$6970000 \times 0.75) - \$862195 = \$4364300$	$\$6361489 + \$4364300 = \$10725790$
4	$(\$6970000 \times 0.85) - \$862195 = \$5061166$	$\$10725790 + \$5061166 = \$15786956$
5	$(\$6970000 \times 0.9) - \$862195 = \$5409599$	$\$15786956 + \$5409599 = \$21196555$

In year 0, the net cash inflow is just the implementation cost. For years 1 to 5, the net cash flow is the savings minus the annual operating cost of the model. The savings is multiplied by different values each year. This is based on the assumption that while the Type 2 model has a 95% accuracy, when implemented to a much larger and complex water distribution system, the savings will not exactly be that high. In the first year, we assume a 60% savings and then it increases to 70% in the second year. As the model is refined each year, the potential savings will increase with better maintenance of the system and understanding of how the model works.

Based on table X above, between year 0 and 1 is when we see a positive cumulative cashflow. So, the payback period is calculated as follows:

$$Payback\ Period = \frac{\$973379.06}{\$3319805} + 1\ year = \ 1.3\ years$$

References

Bakhtawar, B. and Zayed, T., Review of Water Leak Detection and Localization Methods through Hydrophone Technology. *Journal of Pipeline Systems Engineering and Practice*, vol. 12, no. 4, 2021.

Cody, R. A., Tolson, B. A. and Orchard, J., Detecting Leaks in Water Distribution Pipes Using a Deep Autoencoder and Hydroacoustic Spectrograms, *Journal of Computing in Civil Engineering*, vol. 34, no. 2, 2020.

Earth Observatory, Lake Mead keeps Dropping, Available: https://earthobservatory.nasa.gov/images/150111/lake-mead-keeps-dropping, Accessed on March 14, 2023.

GIGABYTE, G493-SB3 (rev. AAP1) | GPU Servers - GIGABYTE Global (n.d.), Available: https://www.gigabyte.com/Enterprise/GPU-Server/G493-SB3-rev-AAP1, Accessed on March 23, 2023.

Government of Canada, Canada's submission to the United Nations Framework Convention on Climate Change, Available: https://www.publications.gc.ca/collections/collection_2021/eccc/En81-4-2019-1-eng.pdf, Accessed on March 14, 2023.

ITM Instruments INC, FLOMEC QS200-30 Insertion Ultrasonic Flowmeter for 3" pipes (n.d.), Available: https://www.itm.com/product/flomec-qs200-30-insertion-ultrasonic-flowmeter-for-3-inch-pipes, Accessed on April 9, 2023.

Job Bank, Wages in the Kitchener–Waterloo–Barrie Region (n.d.), Available: https://www.jobbank.gc.ca/wagereport/location/39328, Accessed on March 23, 2023.

Kammoun, M., Kammoun, A. and Abid, M., Experiments based comparative evaluations of machine learning techniques for leak detection in water distribution systems, *Water Supply*, vol. 22, no. 1, pp. 628–642, 2022.

Kammoun, M., Kammoun, A. and Abid, M, LSTM-AE-WLDL: Unsupervised LSTM Auto-Encoders for Leak Detection and Location in Water Distribution Networks, *Water Resources Management*, vol. 37, no. 2, pp. 731-746, 2023.

Khulief, Y.A., Khalifa, A., Mansour, R. B. and Habib, M.A., Acoustic Detection of Leaks in Water Pipelines Using Measurements inside Pipe, *Journal of Pipeline Systems Engineering and Practice*, vol. 3, no. 2, pp. 47-54, 2012.

Plappally, A.K. and Lienhard V, J.H., Energy requirements for water production, treatment, end use, reclamation, and disposal, Renewable and Sustainable Energy Reviews, vol. 16, no. 7, pp. 4818-4848, 2012.

Punukollu, H., Vasan, A., and Srinivasa Raju, K., Leak detection in water distribution networks using deep learning, *ISH Journal of Hydraulic Engineering*, pp. 1-9, 2022.

Region of Waterloo, Region of Waterloo Corporate Energy Plan 2019-2028, Available: https://www.regionofwaterloo.ca/en/regional-government/resources/Environment/Corporate-Energy-Plan-access.pdf, 2019.

SCG Flowmetrix, Traditional Acoustic Watermain Leak Detection Survey, CITY OF KITCHENER LEAK DETECTION REPORT – 2021.pdf, 2021.

Swain, R. R., Khilar, P.M. and Bhoi, S. K., Underlying and Persistence Fault Diagnosis in Wireless Sensor Networks Using Majority Neighbors Co-ordination Approach, *Wireless Personal Communications*, vol. 111, pp. 763-798, 2020.

Utility rates - City of Kitchener (n.d.), Available: https://www.kitchener.ca/en/taxes-utilities-and-finance/utility-rates.aspx, Accessed on April 1, 2023.

VERIS, Pressure Sensors | Static Pressure Gauges | Veris (n.d.), Available: https://www.veris.com/126864/category/pressure-sensors/static-pressure-gauges, Accessed on April 9, 2023.

Vrachimis, S.G., Kyriakou, M.S., Eliades, D.G. and Polycarpou, M.M., LeakDB: A benchmark dataset for leakage diagnosis in water distribution networks, Available: https://github.com/EpistasisLab/penn-ml-benchmarks, 2018.

Water Scarcity, International Decade for Action “Water for Life” 2005-2015. (n.d.), Available: https://www.un.org/waterforlifedecade/scarcity.shtml, Accessed on April 9, 2023.

WaterWorld, Listening for Leaks (n.d.), Available: https://www.waterworld.com/drinking-water/distribution/article/14235456/2204wwft2, Accessed on April 5, 2023.

Xu, Y., Noy, A., Lin, M., Qian, Q., Li, H. and Jin, R., WeMix: How to Better Utilize Data Augmentation, *arXiv preprint arXiv:2010.01267*, 2020.

Biographies

Hao Wang is an undergraduate in Chemical Engineering (University of Waterloo, CA).
Shiani Raj is an undergraduate in Chemical Engineering (University of Waterloo, CA).
Troy Lewis is an undergraduate in Chemical Engineering (University of Waterloo, CA).
Zhen Ye is an undergraduate in Chemical Engineering (University of Waterloo, CA).
Haitian Zhang is a Ph.D. Candidate in Chemical Engineering (University of Waterloo, CA). Her research focuses on process monitoring and fault diagnosis based on machine learning algorithms in the field of process systems engineering. Prior to her doctoral study, she secured her Bachelor’s and Master’s degrees in Chemical Engineering from Xi’an Jiaotong University. She had two PRES conference papers and three journal papers in her masteral program. During the period of her doctoral study, she also published two papers related to multivariate statistical analysis.
Hamid-Reza Kariminia is a Lecturer in Chemical Engineering at the University of Waterloo. His diverse research area involves investigating sustainable biological solutions to address environmental problems. He has extensive international experience in engineering consultation and worked with various industries including oil and gas, wastewater treatment, chemical manufacturing, and pharmaceutical. He received his BSc and MSc in chemical engineering from Sharif University of Technology, and PhD from Kagoshima University. Prior to joining the University of Waterloo, he served as an Associate Professor at the Sharif University of Technology. His research interests include bio-electrochemical systems, biodiesel production, wastewater treatment, bioremediation, and sustainable production operations.
Ali Elkamel is a Full Professor of Chemical Engineering. He is also cross appointed in Systems Design Engineering. He holds a BSc in Chemical Engineering and BSc in Mathematics from Colorado School of Mines, MSc in Chemical Engineering from the University of Colorado, and PhD in Chemical Engineering from Purdue University. His specific research interests are in computer-aided modeling, optimization, and simulation with applications to energy planning, sustainable operations, and product design. His activities include teaching graduate and undergraduate courses, supervising post doctorate and research associates, and participation in both university and professional societal activities. He is also engaged in initiating and leading academic and industrial teams, establishing international and regional research collaboration programs with industrial partners, national laboratories, and international research institutes. He supervised over 120 graduate students (of which 47 are PhDs) and more than 45 post-doctoral fellows/research associates. He has been funded for several research projects from government and industry. Among his accomplishments are the Research Excellence Award, the Excellence in

Graduate Supervision Award, the Outstanding Faculty Award, and IEOM Awards. He has more than 425 journal articles, 175 proceedings, 50 book chapters, and has been an invited speaker on numerous occasions at academic institutions throughout the world. He is also a co-author of six books.