

A LEAK LOCALISATION SYSTEM IN WATER  
DISTRIBUTED NETWORK USING MACHINE  
LEARNING

PNG WEN HAO

DOCTOR OF PHILOSOPHY (ENGINEERING)

LEE KONG CHIAN FACULTY OF ENGINEERING &  
SCIENCE  
UNIVERSITI TUNKU ABDUL RAHMAN  
APRIL 2022

**A LEAK LOCALISATION SYSTEM IN WATER  
DISTRIBUTED NETWORK USING MACHINE LEARNING**

By

**PNG WEN HAO**

A thesis submitted to the Department of Electrical and Electronics  
Engineering,  
Lee Kong Chian Faculty of Engineering & Science,  
Universiti Tunku Abdul Rahman,  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy (Engineering)  
April 2022

## **ABSTRACT**

### **A LEAK LOCALISATION SYSTEM IN WATER DISTRIBUTED NETWORK USING MACHINE LEARNING**

**Png Wen Hao**

Water crisis has preponderated worldwide due to non-effective resource management and uncontrolled water losses. Leakage appears to be one of the major factors of the water losses. In the past decades, various detection and localisation approaches have been established to solve the problem, acoustic sensing is an effective localisation technique which has been extensively implemented in single pipeline system. However, the conventional acoustic sensing technique faces multiple challenges such as analytics complexity and time-consuming issue in complex piping network system. In this thesis, two leak localisation systems based on machine learning and remote-acoustic sensor network were developed as the solution.

In the first part of thesis, a multi-level analytics framework (MLAF) was formulated for adaptive leak localisation in piping networks. The MLAF analyses the complex spatial acoustic signals from the sensor network and predicts the leak location through multi-level hierarchical analyses and sequential reasoning processes. The system aggregated path analysis, time-correlation location analysis, and machine learning methods to predict the leak location. The processes were handled by an automated flow control to ensure time-effective prediction without needs of human supervision. The performance of MLAF in varying shaped piping networks has been validated by a set of

characterisation tests. The result of a field prediction in a local district metered area (DMA) with excellent root mean square error (RMSE) has further confirmed the feasibility of the MLAF.

In the second part of thesis, a mixed-model deep neural network (MDNN) was designed for alternative leak localisation. The system serves for non-adaptive application in a targeted piping network. The MDNN model comprises multi-layer classification and regression neural networks which was constructed based on Keras module. The MDNN was trained with a set of simulated leak data of the targeted piping network to identify the leak segment and location. Various neural network's hyperparameters such as tensor shape, batch size and sample size were tuned during the training processes to identify the optimal model of prediction. The optimal MDNN model was validated with 2.3% mean absolute percentage error (MAPE) and 0.99 training accuracies. Finally, the optimal MDNN model was implemented to predict the leak location in the local DMA and achieved an excellent result of an average 3.2% location error.

## **ACKNOWLEDGEMENTS**


**My highest gratitude to my supervisor, Dr. Pua Chang Hong and my co-supervisor, Dr. Tham Mau Luen. They have given plenty of constructive suggestions and clear guidance while I was completing my research. A good ethic of research was inculcated by them during the period of my PhD's degree.**

**This material is based upon work performed at University Tunku Abdul Rahman and supported by Grant ONRG/N62909-18-2014, Grant IPSR/RMC/UTARRF/2019-C1/L08, Grant TRGS/1/2016/UTAR/01/2/1, and under collaboration with Ranhill Water Services (RWS) Sdn. Bhd., Malaysia.**

## APPROVAL SHEET

This dissertation/thesis entitled “**A LEAK LOCALISATION SYSTEM IN WATER DISTRIBUTED NETWORK USING MACHINE LEARNING**” was prepared by PNG WEN HAO and submitted as partial fulfillment of the requirements for the degree of Master of Engineering Science at Universiti Tunku Abdul Rahman.

Approved by:



\_\_\_\_\_  
(Pua Chang Hong)

Professor/Supervisor

Department of Electrical and Electronics Engineering

Lee Kong Chian Faculty of Engineering & Science

Universiti Tunku Abdul Rahman

Date: 17/9/21 .....



\_\_\_\_\_  
(Tham Mau Luen)

Professor/Co-supervisor

Department of Electrical and Electronics Engineering

Lee Kong Chian Faculty of Engineering & Science

Universiti Tunku Abdul Rahman

Date: 17/9/2021 .....

**LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE**

**UNIVERSITI TUNKU ABDUL RAHMAN**

Date: 10/9/2021

**SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS**

It is hereby certified that *Png Wen Hao* (ID No: *1903864* ) has completed this final year project/ dissertation/ thesis\* entitled "*A LEAK LOCALISATION SYSTEM IN WATER DISTRIBUTED NETWORK USING MACHINE LEARNING*" under the supervision of Dr. Pua Chang Hong (Supervisor) from the Department of Electrical and Electronic Engineering, Lee Kong Chian Faculty of Engineering and Science, and Dr. Tham Mau Luen (Co-Supervisor)\* from the Department of Electrical and Electronic Engineering, Lee Kong Chian Faculty of Engineering and Science.

I understand that University will upload softcopy of my final year project / dissertation/ thesis\* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,



---

(*Png Wen Hao*)

## DECLARATION

I hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

Name           PNG WEN HAO          .

Date           10/9/2021          .



## LIST OF TABLES

<b>Table</b>		<b>Page</b>
2.1	Summary of existing leak detection methods.	14
2.2	Summary of heterogeneous sensing leak localisation systems in WDN.	19
2.3	Summary of modelling and supplementary analyses.	23
4.1	Localisation results of validation tests 1 and 2.	67
4.2	Localisation results in respects with field measurements and simulated inputs.	73
4.3	Simulated input data and respective targets.	74
4.4	Training results based on tensor shapes 1, 2, 3, and 4.	82
4.5	Training results based on batch sizes 32, 64, and 128.	85
4.6	Training results based on the five sample sizes.	88
4.7	Results of the performance test.	89
4.8	Results of the field prediction.	90

## LIST OF FIGURES

<b>Figures</b>		<b>Page</b>
2.1	Internal and external leak detection methods.	9
2.2	Deployment of the 295 remote noise loggers in the State of Johor (WaterWorld, 2019).	24 26
2.3	Schematics of the typical acoustic-based leak localisation setup (Li et al., 2014).	29
2.4	The TDoA and peak frequency in the CTFS (Foo, et al., 2018).	29
2.5	Dispersion curve of the copper pipeline plotted using a free software package developed in MATLAB by (Seco & Jiménez, 2012).	30
2.6	Multi-directional waves in piping network.	30
3.1	Outline of the MLAF.	33
3.2	Outline of the path analysis and paths storing.	35
3.3	Location analysis based on APA and SPA.	36
3.4	Outline of the backward path analysis and data reorganisation (example case: APA estimation results of correlation between sensors 2 and 4).	37
3.5	General statistical analysis, pruning and pinpointing.	38
3.6	Core points, border points, and noise points (outliers).	40
3.7	DBSCAN clustering process.	40
3.8	Outline of MDNN localisation.	44
3.9	Basic neural network model.	45
3.10	The mixed-model architecture of MDNN.	47
3.11	Overfitting and early stopping (Serveh, et al., 2019).	50

3.12	Dropout of neural network (Mohamed & Hichem, 2018)	52
3.13	Sample case of leak data simulation.	53
3.14	Summary of methodologies 1 and 2.	55
4.1	The 20 emulated piping networks with different network sizes and leak locations.	57
4.2	(a), (b) : First stage evaluations based on the degree of dominance.	59
4.3	(a), (b): First stage evaluations based on the degree of dominance.	59
4.4	(a), (b), (c), (d): Second stage evaluations based on the RMSEs	60
4.5	A quadrilateral DMAs with different lengths ratio.	61
4.6	(a), (b) : First stage evaluations based on the degree of dominance.	62
4.7	(a), (b): First stage evaluations based on the degree of dominance.	62
4.8	(a), (b), (c), (d): Second stage evaluations based on the RMSEs	63
4.9	Emulated DMA 1.	64
4.10	6th level general statistical analysis.	65
4.11	7th level pruning.	65
4.12	8th level pinpointing.	65
4.13	Simulated DMA 2 with non-uniform shape.	66
4.14	6th level general statistical analysis.	66
4.15	7th level pruning.	67
4.16	8th level pinpointing.	67
4.17	Local DMA at Jalan Pelanduk, Taman Scientex, Masai, Pasir Gudang, Johor. (GPS coordinate: 1°30'42.3"N 103°54'57.8"E).	68
4.18	Installation of the hydrophone sensor.	69

4.19	Acoustic signals captured by the six hydrophone sensors.	69
4.20	TDoAs between respective acoustic signals.	70
4.21	(a), (b): 6th level general statistical analysis based on simulated inputs and field measurements.	71
4.22	(a), (b): 7th level pruning based on simulated inputs and field measurements.	72
4.23	(a), (b): 8th level pinpointing based on simulated inputs and field measurements.	72
4.24	Simulated leaks in the targeted DMA	74
4.25	Tensor shape 1.	76
4.26	Tensor shape 2.	77
4.27	Tensor shape 3.	78
4.28	Tensor shape 4.	79
4.29	Training results of tensor shape 1.	80
4.30	Training results of tensor shape 2.	80
4.31	Training results of tensor shape 3.	81
4.32	Training results of tensor shape 4.	81
4.33	Training results based on 32 samples batch size.	83
4.34	Training results based on 64 samples batch size.	83
4.35	Training results based on 128 samples batch size.	84
4.36	Training results based on 3525 sample size.	85
4.37	Training results based on 7050 sample size.	85
4.38	Training results based on 35254 sample size.	86
4.39	Training results based on 70508 sample size.	86
4.40	Training results based on 352540 sample size.	87
4.41	Results of the field prediction.	91

## LIST OF ABVREVIATIONS

ANN	Artificial Neural Network
APA	All-path Analysis
BFS	Breadth-first Search
CNN	Classification Neural Network
CTFS	Cross-time-Frequency Spectrum
DBSCAN	Deep Neural Network
DMA	Density-based Spatial Clustering of Applications with Noise
DNN	District Metered Area
FSR	Force Sensitive Resistor
GIS	Geographic Information System
GPS	Global Positioning System
LCF	Leak Confident Factor
MAPE	Mean Absolute Percentage Error
MCU	Micro Controller Unit
MDNN	Micro-electro-Mechanical System
MEMS	Million litres per day
MLAF	Mixed-model Deep Neural Network
MLD	Multi-level Analytics Framework

NRW	Non-revenue water
RBF-NN	Radial Basis Function Neural Network
ReLU	Rectified Linear Unit
RNN	Regression Neural Network
SPA	Shortest-path Analysis
SSE	Sum of Squares Error
TDNN	Time Difference of Arrival
TDoA	Time-delay Neural Network
UWDM	Urban Water Distribution Management
WDN	Water Distribution Network
WDS	Water Distribution System
WSN	Wireless Sensor Network

## TABLE OF CONTENTS

	<b>Page</b>
<b>ABSTRACT</b>	<b>i</b>
<b>ACKNOWLEDGEMENTS</b>	<b>iii</b>
<b>APPROVAL SHEET</b>	<b>iv</b>
<b>LIST OF TABLES</b>	<b>vii</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF ABVREVIATIONS</b>	<b>xi</b>
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	<b>1</b>
<b>1.1 Background</b>	<b>1</b>
<b>1.2 Problem statements</b>	<b>3</b>
<b>1.3 Aims and objectives</b>	<b>4</b>
<b>1.4 Research Contributions</b>	<b>5</b>
<b>1.5 Thesis outline</b>	<b>6</b>
<b>2 LITERATURE REVIEW</b>	<b>9</b>
<b>2.1 Leak detection methods</b>	<b>9</b>
<b>2.2 Leak localisation systems in WDS</b>	<b>14</b>
2.2.1 Heterogeneous sensing and hardware integrations	14
2.2.2 Modelling and supplementary analyses	19
<b>2.3 Existing leak localisation system in Malaysia</b>	<b>24</b>
<b>2.4 Leak detection based on acoustic sensing</b>	<b>25</b>
2.4.1 Frequency-varying acoustic speed	27
2.4.2 Multi-directional waves issue	30
<b>2.5 Leak detection in oil and gas pipeline</b>	
<b>3 METHODOLOGY</b>	<b>32</b>
<b>3.1 MLAF</b>	<b>32</b>
3.1.1 Path analysis	34
3.1.2 Paths storing	35
3.1.3 Location analysis	36
3.1.4 Backward path analysis and data reorganisation	37
3.1.5 General statistical analysis	38
3.1.6 Pruning	39
3.1.7 Pinpointing	41
<b>3.2 MDNN</b>	<b>43</b>

3.2.1	Feed-forward and back-propagation	45
3.2.2	Mixed-model architecture	47
3.2.3	Overfitting in neural network	50
<b>3.3</b>	<b>Mathematical modelling of the arrival times</b>	<b>52</b>
<b>3.4</b>	<b>Summary</b>	<b>54</b>
<b>4</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>56</b>
<b>4.1</b>	<b>MLAF leak localisation</b>	<b>56</b>
4.1.1	Characterisation based on network sizes	56
4.1.2	Characterisation based on network shapes	60
4.1.3	Validation tests based on non-uniform DMAs	64
4.1.4	Field prediction in the local DMA	68
<b>4.2</b>	<b>MDNN leak localisation</b>	<b>74</b>
4.2.1	Data preparation	74
4.2.2	Training based on different tensor shapes	75
4.2.3	Training based on different batch sizes	82
4.2.4	Training based on different sample sizes	84
4.2.5	Prediction based on the field measurements	89
<b>4.3</b>	<b>Comparisons of the MLAF and MDNN</b>	<b>91</b>
4.3.1	Localisation accuracies	91
4.3.2	Localisation times	92
4.3.3	Flexibilities of application	92
<b>5</b>	<b>CONCLUSION AND RECOMMENDATIONS</b>	<b>93</b>
<b>5.1</b>	<b>Conclusion</b>	<b>93</b>
<b>5.2</b>	<b>Recommendations for future works</b>	<b>94</b>
	<b>REFERENCES</b>	<b>96</b>
	<b>LIST OF PUBLICATIONS</b>	<b>107</b>



## **CHAPTER 1**

### **INTRODUCTION**

In this work, two machine learning based leak localisation systems were developed for avant-garde application using acoustic sensing technology in water distribution network (WDN). The methods were priorly proposed to solve the analytical issues faced by the conventional acoustic-based localisation method in WDN. The first chapter describes the research background, problem statements, and objectives.

#### **1.1 Background**

Piping networks are extensively utilised in water distribution system (WDS) to cater for multipurpose fluid transmissions. Continuous supply of water resource is vital and challenging for ever-growing urban population and consistently expanding service areas. Water supply through the distribution network faced high-level water loss caused by leakage every year.

In the past decade, leakage in WDS had caused up to 60% of non-revenue water (NRW) worldwide (Levenson & Daley, 2010). Malaysia as a developing country in Southeast Asia, had suffered from an average 37.7% NRW loss back in 2005 (MWA, 2017). Soon, a restructuring exercise was initiated by the Malaysian government in 2006 to incentivise the reduction of

NRW, by standardising policies and addressing expenditures in reforming the water supply sector (Teo, 2014). However, the effect was insignificant with 2.5% NRW reduction after 11 years of the reformations. By 2016, Malaysia still suffered from 35.2% NRW nationwide, while severe losses were reported in the States of Kelantan and Perlis, which recorded 67% and 76% NRWs, respectively (Hidir, 2018). In early 2016, Malaysia had lost up to 5,846 million litres per day (MLD) of treated water. The NRW was consistently high with an average daily loss of 5,929 MLD in 2019. The amount was sufficient to cover the water demand in the States of Selangor (3,316 MLD) and Johor (1,320 MLD) for one year (TheSun, 2019).

To deal with the persistently high NRW loss, Malaysian government initiated a National NRW Reduction Programme to reduce the NRW to 31% nationwide by the end of the 11th Malaysia Plan (11MP) (TheSun, 2019). RM1.39 billion was allocated for implementation of the programme in 2018 and to assist the operators in achieving the nationwide NRW target by 2020. The programme was prior to change and replace the old pipelines in some critical regions. Apropos of the NRW reduction programme, water supply companies such as Ranhill SAJ and Air Selangor Sdn Bhd have showed great commitments by experimenting a few remote monitoring systems.

In 2018, a technology trial was implemented by Ranhill SAJ for SMART District Monitored Area in the State of Johor (WWA, 2018). In the initial phase, 295 remote noise loggers were deployed in five local district metered areas (DMAs) for a three-month real-time monitoring. The trial has successfully reduced the net night flow of the one monitored DMA by one-third.

The result of the trail was remarkable with 115 found leaks during the entire monitoring period (WaterWorld, 2019).

Meanwhile, Air Selangor has reduced the NRW of Capital Kuala Lumpur by 1.78% in 2019 through their in-house Smart Operation Control Centre (TheSun, 2019). Data loggers were deployed in 1,601 reservoirs and 731 pump houses to detect the water pressure transient and monitored by SCADA system (SelangorJournal, 2020). The leak detection programme has successful with 108,759 closed leak cases in 2019, compared to 104,033 cases in 2018.

## **1.2 Problem statements**

In the past decades, research in water leakage detection and localisation have been well-established. Acoustic sensing is one of the most common methods and practically implemented by many water industries. Acoustic sensing has advantages over other detection methods in terms of portability, as most devices come with plug-and play feature. The localisation method using acoustic sensing is relatively straight-forward. By correlating two acoustic signals from the adjacent sensors, the leak location can be predicted based on time-distance relation. The two-sensor localisation method was proven feasible by many experiments in the single pipeline system.

However, in practical, WDN exists in form of complex piping network instead of the single pipeline system. The conventional two-sensor localisation system is not effective and time-consuming to localise a leak in a WDN. Moreover, multi-directional waves of leak in the WDN appears to be the biggest

challenge of the conventional localisation technique as the leak-induced acoustic waves tend to propagate through multiple paths to reach the sensors.

Although in most practical cases, acoustic wave tends to survive through the shortest transmission path due to damping property of pipeline, leak location analysis based on the shortest path assumption seems to be oversimplifying the problem. Generalisation of the problem may affect the localisation result to different extents in different piping networks. The effect becomes obvious when the shortest path does not vary significantly from other transmission paths in a piping network. In that case, unknown transmission paths are taken by the acoustic waves to reach the sensors. The conventional leak localisation based on time-distance relation fails to predict the leak location due to unknown path distance. Besides, interference and superposition of waves might further increase the analytics complexity in WDN due to the path uncertainty issue.

### **1.3 Aims and objectives**

This objective of this thesis is to develop the machine learning based leak localisation system incorporates with acoustic sensing for single leak localisation in WDNs. In addressing the problems faced by the conventional acoustic-based leak localisation technique, as discussed in the previous section, four objectives were set wittingly as follows:

1. To formulate a systematic methodology with full analytics consideration of the multi-directional acoustic waves for proper leak localisation in WDNs.

2. To develop an automated control flow system for fluent data flow handling, seamless analytics, and reasoning processes, thereby achieving time-effective leak localisation.
3. To characterise and validate the adaptiveness of the leak localisation system in various piping networks of different network topologies.
4. To develop an alternative leak localisation system based on deep neural networks to predict leak location in a WDN through feature learning of the simulated time differences.

#### **1.4 Research Contributions**

The major contributions of this thesis were summarised as follows:

##### **1. A Multi-Spatial Acoustic-based Leak Localisation Scheme**

Traditional acoustic localisation method based on two-sensor setup is less efficient and heavily dependent on human decision. The sensors are usually installed after pre-localisation examination to determine the best location for installation. The examination methods include pressure monitoring and field tour by humans. By utilizing the acoustic sensor network setup, time-effective localisation was achieved by multi-spatial leak localisation scheme. The multi-spatial acoustic signals from leak in a piping network were simultaneously collected for aggregated analysis, thereby achieving prompt localisation. The multi-spatial leak localisation scheme has advantages over the conventional method in terms of effectiveness and less demand of human forces.

##### **2. An Autonomous and Adaptive Leak Localisation Algorithm**

An automated machine learning algorithm was developed for intelligent and time-effective leak localisation in WDN. The analyses the multi-spatial acoustic data and the subsequent localisation processes are fully automated and without need of human supervision. In conjunction with the high flexibility and portability of the acoustic sensors, the leak localisation algorithm is formulated to be adaptive so that the multi-spatial leak localisation scheme is applicable in varying piping networks. The autonomous and adaptive leak localisation algorithm incorporates the plug-and-play features of acoustic sensors together delivers a favorable and effective localisation for WDN.

### 3. Alternative Leak Localisation Method using Deep Neural Network

In continuing with multi-spatial leak localisation scheme in part 1, a deep neural network (DNN) based leak localisation was developed as another alternative. Different from high portability of 2, the system was formulated to work independently as a fixed monitoring system in a WDN. A multi-level DNN was specifically built and trained to fit the leak localisation problem based on the targeted WDN. Features such as the piping network topology, size of the sensor network, and the placement of sensor within the network are essential in the DNN model's customisation.

## **1.5 Thesis outline**

The first chapter of this thesis provides a brief introduction to the background of the research. The importances of leak detection and localisation

were discussed to solve the alarming NRW losses in both Malaysia and nationwide. A few problems faced by the conventional acoustic-based localisation techniques were shortly presented. Three corresponding objectives were set at the end of this chapter in addressing the previous problems.

Chapter 2 provides a general review on existing leak detection methods in pipeline system and followed by comparison of various leak detection methods in WDN. Subsequently, an in-depth review of acoustic-based leak detection was presented to discuss the limitations and challenges of the methods in WDN.

The detailed methodologies of the two intelligent leak localisation systems were discussed in Chapter 3. Multi-level analytics framework (MLAF) is first presented for adaptive localisation. Subsequently, MDNN is formulated for fixed monitoring system. Lastly, a mathematical modeling was discussed as the possible way of mimicking the real measurement from the piping networks.

In the first part of Chapter 4, the MLAF was characterised and validated based on various piping network of different topologies. A field test was conducted in a local DMA for further verification. In the second part, the optimal model of MDNN was trained and identified through hyperparameter optimisation. The optimal model was tested by a field prediction similar in the first part. The performances of the two systems were compared in terms of localisation accuracy, processing time, flexibility, and the limitations.

Chapter 5 concludes the practicalities and limitations of the MLAF and MDNN based on their performances in Chapter 4. Finally, some

recommendations to the systems were discussed for the future works and possible improvements.



## CHAPTER 2

### LITERATURE REVIEW

In this chapter, existing leak detection methods in pipeline system were generally reviewed. Various leak detection methods in WDN were compared and followed by an in-depth review of acoustic-based leak detection. Limitations and challenges of the method in WDN were summarized at the end of this section.

#### 2.1 Leak detection methods

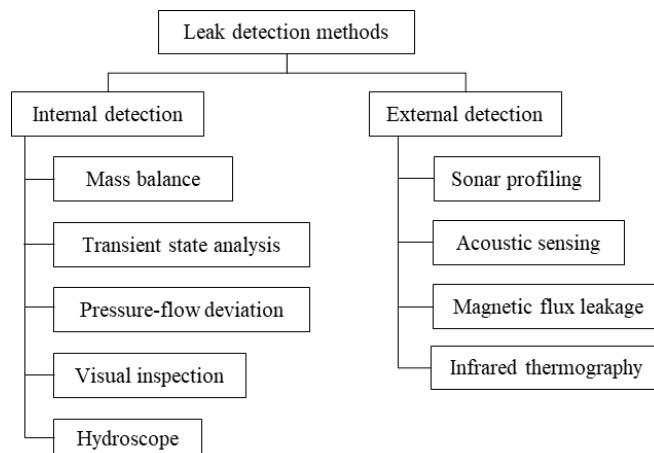


Figure 2.1: Internal and external leak detection methods.

The existing leakage detection methods can be generally classified into two groups: internal and external detection methods (Anon., 1999; Santos, et al.,

2011). Internal leak detection methods detect leak by investigating the internal conditions of pipelines and the physical observables such as water volume and pressure. As shown in Figure 2.1, mass balance, transient state analysis, pressure-flow deviation, visual inspection, hydroscope, and permeability scan are the common internal detection methods. In general, these methods shown effective results in detecting leak and the characterizing the leak features. Internal scanning techniques have achieved accurate leak localisation in many practical inspections. However, leak localisation is impossible through investigations of water volume and pressure (Ghazali, et al., 2010; Makeen, et al., 2014; Abdulshaheed, et al., 2017).

Mass balance is one of the most common internal detection methods used in water audits and assessments (Wang, et al., 2001). The method is simple and straight-forward by measuring the difference between the inlet volume and the outlet volume. Leakage can be easily identified by calculating the change of mass. However, this method fails to detect the exact leak location in pipelines.

Transient state analysis is another internal detection method which relies on damping of the transient pressure waves in a pipeline (Makeen, et al., 2014). During leakage, flow disturbance at leak hole generates transient pressure waves. The transients pressure waves decay as they propagate through the pipeline due to attenuation. Instruments such as strain gauge pressure transducer and Piezotronics pressure sensor are commonly used for transient state analysis. The method detected leak by monitoring water pressure at single points along the pipeline, which is also called pressure point analysis (Afifi, et al., 2011).

Pressure-flow deviation is a model-based leak detection method which detects and estimates the leak location by analysing the deviation of water pressures and flowrates (Wang, et al., 2001). The method requires installation of pressure and flow sensors in strategic points of a pipeline. The detection was carried out by matching the actual measurements to the simulation model measurements. Leak location was identified by measuring the deviation of the measurements from the expected values.

Closed-circuit television (CCTV) and laser scan are the popular visual inspection techniques. Both techniques require installation of camera or scanning device inside a pipeline to examine the internal conditions of the pipeline in finding leak (Zangenhmadar & Moselhi, 2014). Laser scan is a costly inspection system which caters for in-depth monitoring includes finding minor deflections and the corrosion. Comparatively, CCTV is a cheaper alternative which monitors the internal condition of a pipeline in real-time and often operates in conjunction with image processing techniques.

Hydroscope is an internal inspection method based on the principle of eddy current (Makar & Mcdonald, 1996; Rizzo, 2010). Time-varying magnetic field was generated by magnetic coil in the hydroscope. Different characteristics of the pipeline such as wall thickness, discontinuities, corrosion, and crack induce different degrees opposing magnetic field to the magnetic coil. Thus, the wall of pipeline can be examined by measuring the changes of impedance changes. However, the method is limited to pipeline of electrically conductive materials.

On the other hand, external detection methods such as sonar profiling, visual inspection, acoustic sensing, magnetic flux leakage, and infrared thermography have advantage over the internal detection methods in terms of flexibility and ease of implementation (Hunaidi & Wang, 2004). The devices are usually plug-and-play and portable. The sensing set up can be easily installed and operated from the exterior of pipeline. Besides, pinpointing of leak locations are possible and proven with high accuracies by these methods (Sun & Li, 2010; Maninder, et al., 2010; Butterfield, et al., 2017).

The sonar profiling system consists of a movable scanner device (transmitter) to send sound waves through the pipeline and a receiver to measure the reflected waves caused by discontinuity on the pipeline (Zangenehmada, 2014). Operational frequency of the sonar profiling system is adjustable for different pipelines and detection proposes. In general, sonar profiling system under higher frequencies resulted a better resolution, whereas low frequencies operation has a higher sound wave penetration.

Acoustic sensing is another external detection method which is commonly implemented in practical inspection by many water services. In the previous study, leak detection based on leak-induced acoustic wave has been validated experimentally in a closed-loop pipeline system (Law, et al., 2018; Png, et al., 2018). Devices such as accelerometers, hydrophones, and noise loggers were utilised to capture the leak-induced acoustic waves in detecting leakage (Eiswirth, et al., 2001; Kleij & Stephenson, 2002; Ravichandran, et al., 2021). Furthermore, the method was commonly adopting time-correlation technique in localising leak (Li, et al., 2014; Foo, et al., 2018).

Magnetic flux leakage method has shown high feasibility in detecting minor defections and corrosion in pipelines (Nestleroth & Bubenik, 1999; Mukhopadhyay & Srivastava, 2000; Afzal & Udpa, 2002). During inspection, electro-magnetic pig was mounted onto the exterior wall of a pipeline. Magnetic flux generated by the pig travels in the axial direction of the pipeline. The density of magnetic flux was measured by the flux-sensitive sensor during the surface scanning. Leak can be detected by measuring the variation of magnetic flux in a pipeline as reduction of wall thickness and discontinuity in pipeline may cause drop in flux density.

Infrared thermography detects water leakage by measuring the ambient soil temperature of a pipeline (Wang, et al., 2001). Thermographic camera and scanner were commonly used to capture the thermographic images of soil. The method is effective in locating underground leak, as the ambient temperature of soil around the pipeline drops when leak happens. However, the method was limited by the depth of underground leak. The results also affected by the surface conditions and moisture of the ambient ground. Lastly, Table 2.1 summaries the capabilities of the existing leak detection methods in terms of detection and localisation.

Table 2.1: Summary of existing leak detection methods.

Methods	Detection	Localisation
Mass balance	✓	
Transient state analysis	✓	
Pressure-flow deviation	✓	✓
Visual Inspection	✓	✓
Hydroscope	✓	✓
Sonar profiling	✓	✓
Acoustic sensing	✓	✓
Magnetic flux leakage	✓	✓
Infrared thermography	✓	✓

## 2.2 Leak localisation systems in WDS

In the previous section, some common and reliable leak detection methods in single pipeline system were discussed. However, most of the methods are having difficulty to locate leak in WDS. Complexities in both field inspections and post measurement analyses are the major challenges faced by the conventional methods in WDN. Additional manpower, hardware system integrations, and supplementary analyses may be required to solve the complex of localisation in WDN.

### 2.2.1 Heterogeneous sensing and hardware integrations

Heterogeneous sensing and hardware integrations were commonly implemented to solve the localisation problems in WDS. In this section, a few groups of researchers have integrated multiple sensing devices such as pressure,

flow, temperature, and acoustic sensors to deliver a thorough monitoring system.

A distributed ad-hoc wireless sensor network (WSN) was demonstrated by Agathokleous et al. for urban water distribution management (UWDM) in 2015. The monitoring system was experimented in a controlled laboratory environment with different cases of leakage simulations. The WSN was supported by an integrated management system which combines the analytic, geographic information system (GIS), and neuro-fuzzy decision support systems to provide a real-time data acquisition and signals processing (Agathokleous, et al., 2015). The method had achieved reliable results in detecting leak within the WDNs, by monitoring the soil and air conditions such as the temperature and humidity, as well as the water pressure and flow. However, the preliminary findings showed that leak detecting leak based on the moisture readings is less effective compared to the pressure sensors and noise loggers. Furthermore, an entropy-based sensor placement was proposed as a strategic planning to optimise the sensor placement so that the least possible number of sensors is required to achieve equivalent monitoring. A mathematical model incorporated with graph theory and GIS-based spatial analysis is utilised in optimizing for the placement of the pressure sensors and the noise loggers within the WDN. The optimisation was carried out by subcategorised the noise loggers in each regional DMAs of the WDN after the placement of the pressure sensors. In overall, the research focused on multiple devices monitoring system with optimisation of sensors distribution for minimum power consumption and quick converging leak detection. However, leak localisation and field application were not properly addressed in the research.

PipeTECT is another heterogenous sensing system incorporating micro-electro-mechanical systems (MEMS)-based WSN proposed by Shinozuka et al in 2010. The system was experimented on a lab-scaled Polyvinyl chloride (PVC) piping network with reliable result in detecting multiple leaks (Shinozuka, et al., 2010). The WSN consists of eight Gophers (sensing units), four pressure gauges, and one Roocas (wireless communication unit). Both wired and wireless interfaces were used in the node connection for data transmission and communication uplink. Each Gopher sensing node is equipped MEMS-based accelerometer in X, Y, and Z directions, in order to measure the axial vibration on the pipe surface and determine the change in water pressure caused by artificial leaks. Their preliminary results shown significant acceleration changes in sensing nodes located nearer to the leak locations. In the experiment, the peak amplitudes of accelerations from the WSN were used as the main parameter in detecting leak. The peak accelerations of each sensor node were plotted in contour map of the piping network. The leak locations were identified though the intensity levels of the contour map. However, the positions of leak were known regionally by referring to the nearest sensing units, without a detailed numerical value of the exact leak location.

Sun et al. introduced a magnetic induction-based wireless sensor network architecture for underground pipeline monitoring (MISE-PIPE) to detect and localise leakages in underground pipelines. MISE-PIPE integrated internal and external detection methods such as pressure sensors, acoustic sensors, and soil property sensors in the monitoring system (Sun, et al., 2011). A three-phase detection strategic framework was adopted with implementation of MISE-PIPE system in a straight pipeline section in a WDN for effective



monitoring. The 1<sup>st</sup> phase detection was carried out through measurement of the transient event by the pressure sensors. Anomalies in measurement triggered response in 2<sup>nd</sup> phase, where requests were sent by the pressure sensors to the soil property sensors. The soil property sensors in the suspicious areas were activated from the sleep mode and followed by initiation the measurements. The soil properties measurements were aggregated with the pressure measurements and transmitted to the processing hubs above the ground. In 3<sup>rd</sup> phases detection, the processing hubs further examine possibility of leaks in the suspicious areas based on the aggregated measurements from underground. Leakage and the position were reported to the remote administration centre upon confirmation of detection. MISE-PIPE has delivered a full and thorough detection strategic framework for underground pipeline system. However, the proposed method was only experimented in a straight pipeline section in a WDN. Leak localisation results and the accuracies were also not discussed properly in the work.

In 2014, Sadeghioon et al. proposed another smart WSD system to detect underground leak by integrating force sensitive resistor (FSR) based pressure sensors and temperature sensors for internal and external detection. The detection of leak was based on aggregated measurements of the relative indirect pressure changes and the soil temperature (Sadeghioon, et al., 2014). The sensing system consists of four main operating parts: a sensor unit, micro controller unit (MCU), transmission unit, and power management unit. Each sensor unit was equipped with two temperature sensors and one FSR based pressure sensor. In field experiment, the two temperature sensors were installed on the pipe wall and the ambient soil, to measure the pipe wall temperature and

the ambient temperature, respectively. The MCU was attached to the sensor unit for purposes of data acquisition and processing. The measurements data collected by the MCU was sent to other sensor unit via radio frequency transmission. Every four or five sensor units made up one aggregated sensing unit where the signals were handled by one master node. The master unit was capable in transmitting the aggregated data to the cloud through internet connection. Lastly, the power management unit was responsible for effective power consumption through conversion between sleep mode and the active mode of the MCU and transmission unit. The WSN system was experimented in a leak testing facility and showed reliable results in leak detection.

In the end of this section, Table 2.2 summarises all the heterogeneous sensing methods and hardware integration systems discussed previously for comparison. Most of the methods have achieved effective leak detection but not capable of localising the position of leak in the WDN.

Table 2.2: Summary of heterogeneous sensing leak localisation systems in WDN.

Systems	Sensing principles	Limitations
ad-hoc WSN (Agathokleous, et al., 2015)	<ul style="list-style-type: none"> <li>• Temperature</li> <li>• Humidity</li> <li>• Water pressure and flow</li> </ul>	<ul style="list-style-type: none"> <li>• Leak localisation was not achieved</li> <li>• Pipe access was limited in a controlled laboratory environment</li> </ul>
PipeTECT (Shinozuka, et al., 2010)	<ul style="list-style-type: none"> <li>• Acoustic</li> <li>• Pressure</li> </ul>	<ul style="list-style-type: none"> <li>• Leak localisation was not achieved</li> <li>• Pipe access was limited in a controlled laboratory environment</li> </ul>
MISE-PIPE (Sun, et al., 2011)	<ul style="list-style-type: none"> <li>• Water pressure</li> <li>• Acoustic</li> <li>• Soil property</li> </ul>	<ul style="list-style-type: none"> <li>• Leak localisation was claimed without a proper discussion on the accuracies</li> <li>• Pipe access was limited in a straight pipeline section of a WDN</li> </ul>
FSR-based WSN (Sadeghioon, et al., 2014)	<ul style="list-style-type: none"> <li>• Water pressure</li> <li>• Temperature</li> </ul>	<ul style="list-style-type: none"> <li>• Leak localisation was not achieved</li> <li>• Pipe access was limited in a leak testing facility</li> </ul>

## 2.2.2 Modelling and supplementary analyses

Modelling and various supplementary analyses based on machine learning are becoming popular in solving complex problems recent years. EPANET is a prevailing hydraulic modelling software in WDS. The software was used by many researchers in approximating leak location based on fluid dynamics formulation such as mass conservation equation at pipe junctions and

the energy conservation equation around the pseudo-loops of the piping network (Ferziger & Peric, 2002; Bakker, et al., 2003).

In 2012, Wang et al. presented a model-based leak detection method by monitoring the pressure heads in a simulated network model. The network model was simulated using EPANET based on few parameters such as the location and extend of the damage in WSN (Wang, et al., 2012). During the test, a leak was simulated by taps revolving on the pipe up to the 45 degrees clockwise. A statistical prediction models were developed based on the error estimation of the historical pipeline failure data. Genetic algorithm was used to optimise the model in predicting future breakage by identifying the failure patterns. The model has successfully estimated the leakage location and sizes in the small pipelines. However, the estimated leakages in real pipelines were usually smaller than the measured amounts, as the model was built based on the virtual network, where the deviations of the inlet and outlet pressures due to water consumption and supply in the real piping network were ignored in the modelling.

Besides, EPANET is capable to model the WDS by computing the pressures and flows through the demand-driven approach (DDA). The DDA assumes that the water demand is always fulfilled a WSD regardless of the existing pressure, resultant inaccurate estimations in pressure-deficient conditions. In addressing to this problem, Muranho et al. introduced an EPANET extension to compute the pressure–demand/leakage relationship. A pump scheduling was defined using WaterNetGen to effectively reduce the operational costs and the background leakage (Muranho, et al., 2014). However, the pressure-driven solver often incurred numerical instability to the

system loop and caused convergence problems. Therefore, Muranho et al. proposed the relaxation coefficients in the pressure–demand/leakage equation and a smoothing technique for a differentiable pressure–demand/leakage relationship to improve the convergence of the iterative process. The smoothing technique has resulted improvements in most of the simulations but did not promise all-time convergence in every case.

Aside from modelling, soft computing and deep learning have becoming popular in solving complex problems which are unable to be solved by the conventional computing and modelling techniques. Of late, artificial neural network (ANN) have gained high attention in wide ranges of applications. ANN as a mimicry of biological neural system, is effective in many applications such as pattern recognizing and learning through a series of known examples. Without being explicitly programmed, the ANNs are able to solve the complex problems by learning from the large sample of observables (Jafar, et al., 2010). Leakage predictions in WDNs based on the ANN methodologies have been proven feasible by many researchers through analysis of the statistical, environmental, and operational factors.

In 2003, Caputo et al. made an early attempt to implement the ANN in leak localisation for a piping network. Samples of leak and non-leak data were generated through inverse problem of the correlate effects between pressure and flowrate conditions within the network’s boundary conditions (Caputo & Palegagge, 2003). Different sets of data characterizing several states including the perturbations incurred by leaks were used in the training. By recognizing the transient patterns among the data, the ANN was trained to classify the leak data and identify the leak location. Lastly, the trained model of ANNs were

applied for a case study to investigate a leaking pipeline network. The investigation result shown an estimation accuracy within 50–100m of the actual leak location.

Another application of the ANNs in leak detection was effected by Mounce et al. in 2007. Leak identification was achieved by data classification based on the hydraulic parameters including the water flow and pressure in a treated WDS (Mounce & Machell, 2007). A static and time delay neural networks were used to classify the time series pattern for comparison. The static neural network converts a temporal sequence into the static spatial pattern whereas the time-delay neural network (TDNN) stores the past data as short-term memory, allowing the ANN to learn the relationships over time and makes the network dynamic. Both networks were able to recognise the patterns of leaks during the training process. However, the results showed TDNN outperformed the static network, in terms of the mean square error (MSE) of both training and validation test. In burst detection, the TDNN successfully identified 75% of the cases whereas the static neural network only achieved 4%. Additional performance test was conducted based on the experimental data in a DMA of a UK water company where the pipeline bursts were simulated by the opening the hydrant. The estimation of pipeline burst was excellent with accuracy of 98.33%.

In 2014, Salam et al. demonstrated an on-line monitoring system for leak detection in WDS based on the Radial Basis Function neural network (RBF-NN). The RBF-NN consisted of three network layers: an input layer, hidden layer, and an output layer with two neurons. The number of neurons in the hidden layer was determined by trial and error, within the range of maximum

number of neurons which is equal to the number of junctions in the network system (Salam, et al., 2014). The training was conducted based on the simulated data to optimise the parameters such as the smallest training error and the learning rate. Both input data and output data such as the pressure at each junction, magnitude and location of leakage were simulated using EPANET for the network training. The magnitude of simulated leak was manipulated by changing the emitter coefficients with control of leak sizes between 0.01l/s and 0.6l/s. The RBF-NN has achieved 98% prediction accuracy in the magnitude and location of leakage.

Table 2.3: Summary of modelling and supplementary analyses.

Methods	Parameters	Limitations
EPANET modelling (Wang, et al., 2012)	• Mass conservation	• Only satisfied the engineering applications in the small pipelines
EPANET modelling with pressure-driven analysis (Muranho, et al., 2014)	• Mass conservation	• Did not promised an all-time convergence in every case
ANN (Caputo & Palegagge, 2003)	• Water pressure and flow	• Location accuracy achieved within 50–100m
ANNs with static and TDNN (Mounce & Machell, 2007)	• Water pressure and flow	• Leak localisation was not achieved (only detection)
RBF-NN (Salam, et al., 2014)	• Water pressure	• Leak localisation was not achieved (only detection) • Test was performed based on simulated data.

### 2.3 Existing leak localisation system in Malaysia

Of late, Ranhill SAJ performed a technology trial of the SMART District Monitored Area in the State of Johor (WWA, 2018). In the initial phase, 295 remote noise loggers were implemented for real-time monitoring programme and successfully reduced the net night flow in five local DMAs by one-third. During the entire monitoring period, a total number of 115 leaks were found and closed.

The leak monitoring programme continued for an extensive coverage in the multiple areas of the State of Johor. Each logger was equipped with global positioning system (GPS) communication module for daily data transmission and sustained monitoring. Primayer provided the company a cloud-based data collection platform, PrimeWeb to check on the leakage data in all monitoring areas, but only with limited accessibility to it. Figure 2.2 shows the leak confident factor (LCF) of the initial phase of monitoring programme, where the red markers indicate the locations with high possible of leak.

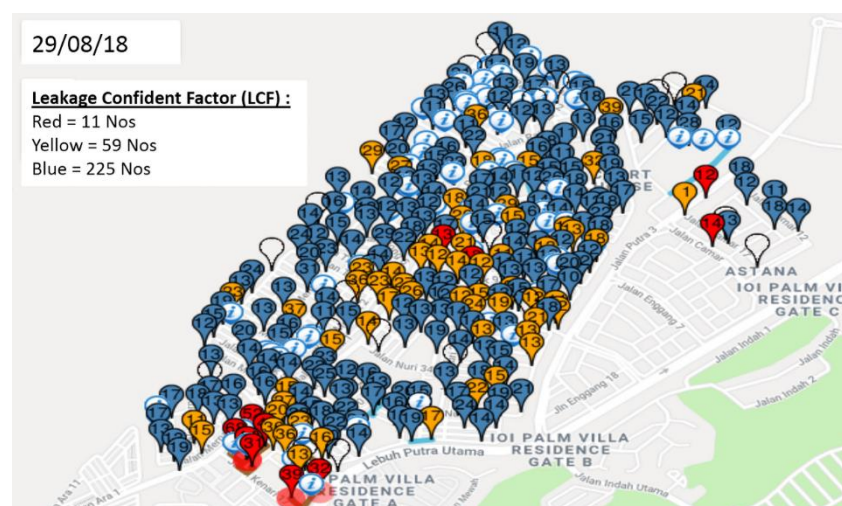


Figure 2.2: Deployment of the 295 remote noise loggers in the State of Johor (WaterWorld, 2019).



A field visit was conducted by me and my research group on 15<sup>th</sup> -17<sup>th</sup> July 2019 to several sites listed in the SMART monitoring area. The tour was guided by a technical crew to provide us detailed demonstration of the monitoring procedures. The early stage of leak detection was done by examining the regions of high LCF through the monitoring system. Subsequently, technical crews were sent to the suspected region for further inspection. Leak around the vicinity of the anomaly region was commonly confirmed by traditional methods such as visual inspection and metal rod listening.

Recently, the inspection teams have started to utilise field detection loggers to localise leak. A set of loggers (usually 2 to 6) were installed on the wall the underground pipe valves to measure the acoustic signal and the collected data will be sent to the local PC for further localisation processes. Conventional cross-correlation method was used to localise the leak. However, manually correlate the multiple signals from the loggers was ineffective. Numerous of trial-and-errors were required to complete the cross-correlation of all pairs of signals from every logger. In some of the tries, the leak location might be undetermined as the leak was out of bound between the two correlated loggers. Therefore, the leak localisation based on trial-and-error correlation was the major challenge for the inspection method.

#### **2.4 Leak detection based on acoustic sensing**

Acoustic sensing has advantages over other leak detection methods in terms of ease of implementation and the capability of localising leak. Besides,

a sole acoustic sensing system has a lower implementation and maintenance costs compare to the heterogenous sensing systems.

During leakage, turbulent and water outlet at the leak spot cause acoustic vibration in pipeline. The frequency of vibration is varied by pipeline material, water flow pressure and the size of leak (Hunaidi, et al., 2004). The leak-induced acoustic vibration is constantly generated by the turbulence in the leaking pipeline. It can be observed as a new emerging spike in the frequency spectrum, where the frequency is distinct from the normal flow-induced acoustic wave. The detection method by differentiating the anomaly flow based on frequency variation has been verified by previous experiments (Khulief & Khalifa, 2012; Png, et al., 2018).

The effectiveness of leak localisation based on the time-correlation principle has been generally proven in the single pipeline systems (Maninder, et al., 2010; Li, et al., 2018; Foo, et al., 2018). Figure 2.3 shows the conventional acoustic-based leak localisation setup in the single pipeline system, where two individual sensors are mounted on exterior of pipeline at two sides of a suspected leakage.

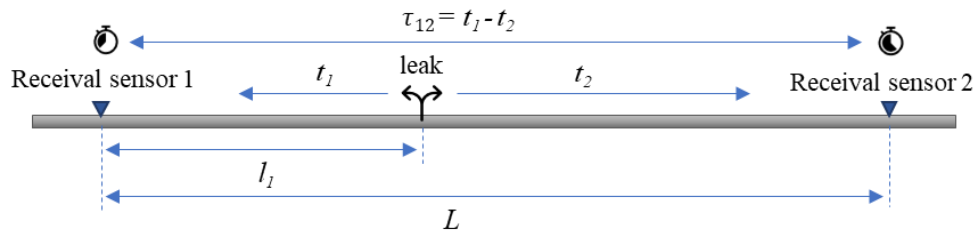


Figure 2.3: The conventional acoustic-based leak localisation setup (Li et al., 2014).

During the leak event, the leak-induced acoustic waves originate from the leak site and propagates along the pipeline in two directions. The sensors

receive the signals of waves at different times, due to difference in the spatial distances. The two corresponding signals can be mathematically modelled as:

$$\begin{cases} x_1(t) = s(t) + n_1(t) \\ x_2(t) = \delta s(t - \tau_1) + n_2(t) \end{cases} \quad (1)$$

where  $x_1(t)$  and  $x_2(t)$  are the received signals by sensors 1 and 2, respectively.  $s(t)$  is the signal from leak with the zero-mean random noises possibly due to ambient interferences,  $n_1(t)$  and  $n_2(t)$ .  $\delta$  is the attenuation factor due to the acoustic path difference and  $\tau_{12}$  is the time delay between the two acoustic signals, which is also known as the time difference of arrival (TDoA).

$$R_{12}(\tau) = E[x_1(t)x_2(t - \tau)] \quad (2)$$

$$\tau_{12} = \arg \max R_{x_1 x_2}(\tau) \quad (3)$$

The TDoA,  $\tau_{12}$  can be determined from equation (3) through the argument of maximum cross-correlation function  $R_{12}(\tau)$ , where  $E$  is the expectation operator of the corresponding time function. Then, the leak location can be estimated from equation (4).

$$l_1 = \frac{L + v\tau_{12}}{2} \quad (4)$$

Where  $v$  is the propagation speed of the acoustic vibration in the given pipeline,  $L$  is the distance between two corresponding sensors and  $l_1$  is the location of leak relative to sensor 1.

#### 2.4.1 Frequency-varying acoustic speed

The propagation speed of the acoustic wave is another important key parameter in localising leak. In many applications, the propagation speed is

assumed to be a constant in pipeline. However, the acoustic speed is highly dependent on the material of pipeline, flow pressure and magnitude of leak. In fact, the acoustic speed varies as a function of the frequency where the acoustic speed-frequency relationship is governed by the dispersive curve of the dominated leakage-induced guided wave mode. In solving the frequency-varying acoustic speed issue, Foo et al. proposed an acoustic speed identification method based on cross-time-frequency spectrum (CTFS) and dispersion curve. The derivation was based on the Fourier transform of the generalised instantaneous cross-correlation function,  $\chi_{x,y}$  with respect to time (Li, et al., 2014), thereby generating the CTFS equation in delay-frequency domain:

$$C_{x,y}(\tau, \omega) = \int \chi'_{x,y}(t, \tau) e^{-i\omega t} dt \quad (5)$$

$$[\omega_0, D] = \arg \max C_{x,y}(\tau, \omega) \quad (6)$$

The TDoA,  $D$  and peak frequency,  $\omega_0$  were then obtained from the CTFS as in equation (6). Figure 2.4 shows the TDoA and peak frequency obtained based on the acoustic data collected from a leaking gas pipeline. With the TDoA and peak frequency, the acoustic speed of the acoustic wave in the investigating pipeline was then determined from the dispersion curve as in Figure 2.5. Lastly, the acoustic wave at the peak frequency,  $v(\omega_0)$  was substituted into equation (4) to estimate the leak location.

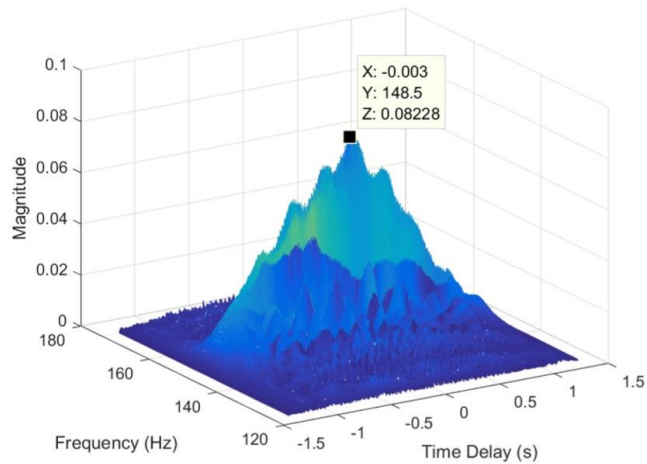


Figure 2.4: The TDoA and peak frequency in the CTFS (Foo, et al., 2018).

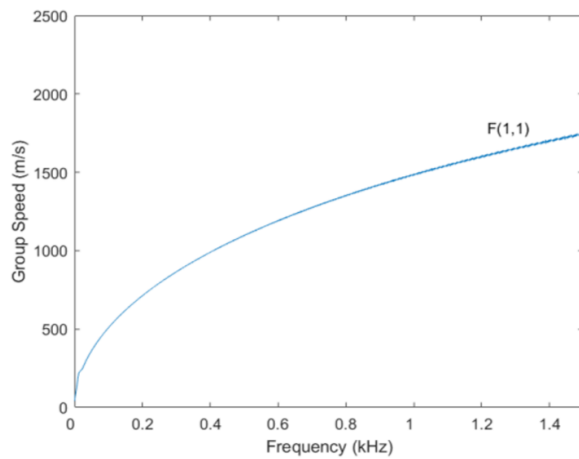


Figure 2.5: Dispersion curve of the copper pipeline plotted using a free software package developed in MATLAB by (Seco & Jiménez, 2012).

Foo et al. have presented a complete and deliberate analytic method to determine the speed of acoustic wave for leak localisation. The method based on the dominant frequency of leak-induced acoustic wave and the material of the pipeline is practical in the real application of leak localisation. The experimental result in a single pipeline leak testing facility was excellent with high localisation accuracy.

### 2.4.2 Multi-directional waves issue

Acoustic sensing incorporated with time-correlation method has been commonly implemented in both research and field application to localise leak in single pipeline system. However, multi-directional waves appear to be the major issue encountered by the conventional method in the piping network. During leakage, acoustic waves originate from the source of leak tend to propagate in various directions throughout the piping network. Figure 2.6 depicts the multi-directional waves propagation in the piping network. The leak-induced acoustic waves are possibly reaching the sensor through various paths. The leak-induced acoustic waves from different paths also tend to interfere with each other and form a super-positioned wave. The phenomenon causes uncertainty in the transmission distance between sensors 1 and 2, therefore leads to confusion of  $L$  in equation (4).

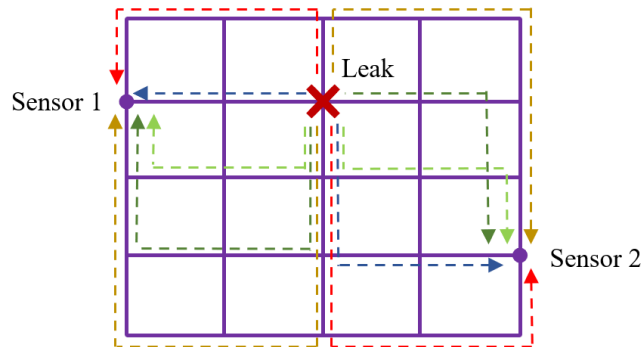


Figure 2.6: Multi-directional waves in piping network.

Although the leak signal tends to survive through the shortest transmission path due to damping property of pipeline in most practical cases, the location analysis solely based on the shortest distance seems oversimplifies the waves transmission problem. Furthermore, the shortest path assumption fails to support the cases that the shortest path does not vary significantly from

other transmission paths in a piping network. Without knowing the exact transmission path taken by leak signal, manual time-correlation method based on trial and error is time-consuming. The leak localisation process takes even longer when involves large number of sensors. Besides, multiple correlation sets of different sensor pairs with unknown transmission paths will eventually leads to an inconclusive localisation result.

For these reasons, two intelligent leak localisation systems are formulated in Chapter 3, to address the issues faced by the conventional time-correlation method and to deliver a time-effective and accurate leak localisation in WDN.

## **2.5 Leak detection methods in oil and gas pipelines**

Leakage in oil and gas pipelines are commonly caused by the aging process. Corrosion and external wear are the major factor of pipeline aging. The process of aging usually takes 15–20 years in a stable operating pipeline (Clair & Sinha, 2014). Oil and gas pipelines have relatively higher pressure than water pipelines (Dong & Yu, 2005). Besides, risk of leakage in oil and gas pipelines is considerably critical with its medium. The leak detection methods in oil and gas pipeline are generally similar with water pipelines. Lidar system, thermal imaging, spectral imaging, acoustic emission, ultrasonic guided wave method, and sonar system are extensively used for leakage detection in the oil and gas pipeline. Due to the high-risk property of oil and gas pipelines, most of the applications focus on corrosion detection to prevent leak, rather than finding the leak. (Lu, et al., 2020)

## **CHAPTER 3**

### **METHODOLOGY**

In this chapter, the detailed methodologies of the MLAF and MDNN were deliberately presented. A mathematical modeling of the arrival times was introduced to mimic the real measurement of leak in the virtual environment. The modeling served as an alternative to simulate leak data for the characterisation tests of MLAF and the MDNN training.

#### **3.1 MLAF**

the MLAF was formulated for an adaptive and time-effective leak localisation in all piping networks. By analyzing the multi-spatial acoustic signal from a sensor network, the MLAF automatically predicts and concludes a leak location. To ensure fluent and seamless localisation process, an automated control flow system was adopted. The overall outline of the MLAF was summarised in Figure 3.1. The MLAF consists of eight hierarchical analytics and reasoning processes, each process was presented and discussed detailly in the following sections respectively.



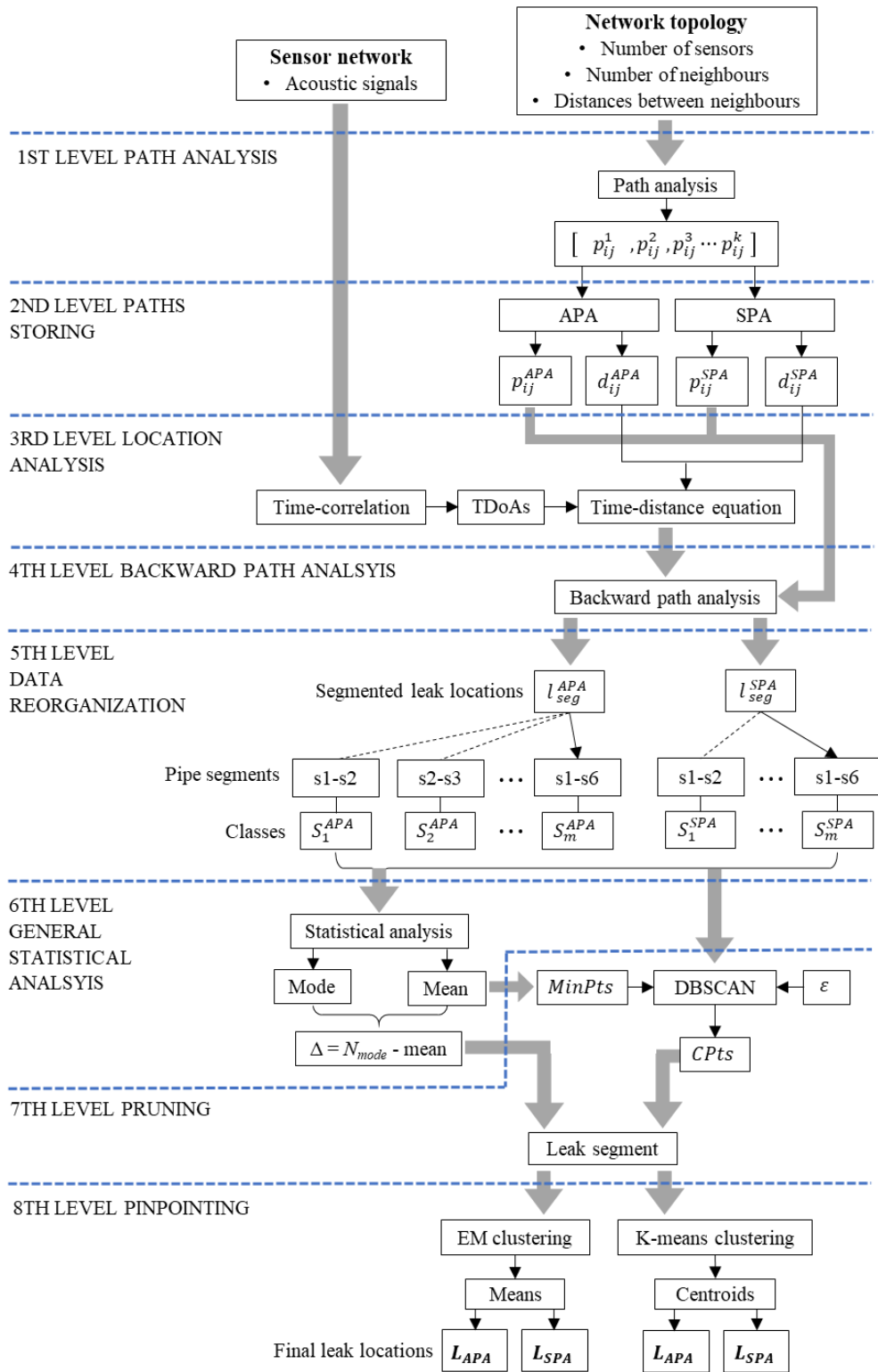


Figure 3.1: The eight hierarchical analytics levels of the MLAF.

In 1<sup>st</sup> level path analysis, all transmission paths between each sensor available in the piping network were identified and stored into individual matrix by 2<sup>nd</sup> level paths storing. The process was followed by location analysis to

determine leak locations based on each path in 3<sup>rd</sup> level. The calculated leak locations were then traced back and reorganised in subsequent two levels. In the last 3 levels, the leak locations were filtered, pruned, and pinpointed through various statistical analytics methods. The detail operation and method of each analytics level will be clearly discussed in the following sections.

### 3.1.1 Path analysis

Path analysis is widely utilised in many routing applications such as GPS, network routing and computer graphics application (Zafar, 2016). Route searching at the lowest cost is a very common application in many shortest path problems. In MLAF, the path analysis was modified for a continuing search of  $k$  numbers of subsequent shortest paths from the source sensor  $i$  to the target sensor  $j$ ,  $p_{ij}^k$ , where  $i$  and  $j = 1, 2, 3, \dots, N$ ,  $N$  is the total number of sensors in a sensor network,  $k = 1, 2, 3, \dots, n$ ,  $n$  is the total number of shortest paths predefined by the user. The subsequent shortest paths always fulfil the criteria such that  $\{ p_{ij}^k \mid d_{ij}^k > d_{ij}^{k-1} \}$  in the iterative search. The initiation of 1<sup>st</sup> level path analysis was based on network topological inputs such as the number of sensors, number of respective neighbours, and distance between respective neighbours. The subsequent  $k$  shortest paths were searched through an iterative breadth-first search (BFS) algorithm. The algorithm continued to explore the piping network until the  $n$  numbers of shortest paths have been thoroughly searched. The searching algorithm then moved on to search for the paths between the next  $i$  and  $j$  sensor pair.

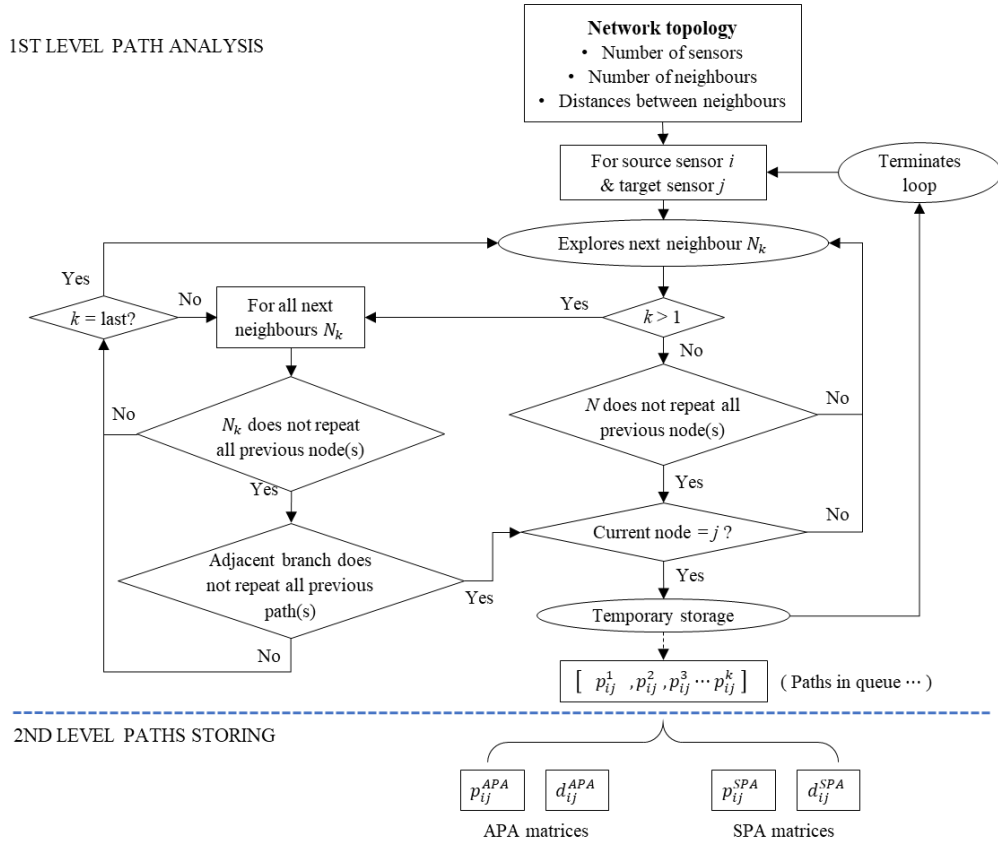


Figure 3.2: Outline of the path analysis and paths storing.

### 3.1.2 Paths storing

The process was followed by the 2<sup>nd</sup> level paths storing upon completion of the 1<sup>st</sup> level path analysis. All the searched paths were stored into two separated matrices, the all-path analysis (APA) matrices and the shortest-path analysis (SPA) matrices. The shortest paths were stored in the  $p_{ij}^{SPA}$  matrix while all the  $n$  paths are stored in the  $p_{ij}^{APA}$  matrix. The corresponding distances were stored in the  $d_{ij}^{APA}$  and  $d_{ij}^{SPA}$  matrices, respectively. The separated storing system based on APA and SPA ensured two distinct flows of hierarchical analyses: one solely based on the shortest paths, another based on multiple transmission paths (up to  $n$  subsequent shortest paths). The SPA retained the

conventional assumption that the shortest path will survive through the transmission due to damping property of pipeline. The APA on the other hand, prevented oversimplification of the wave transmission complexity by considering multiple transmission paths of the leak signal.

### 3.1.3 Location analysis

The 3<sup>rd</sup> level location analysis focused on the analysis of the multi-spatial acoustic signals from the sensor network. At this stage, TDoAs between each pair of signals were determined through a recursive time-correlation analysis. Leak locations,  $l_i$  were estimated based on three parameters,  $\tau_{ij}$ ,  $d_{ij}^{APA}$  and  $d_{ij}^{SPA}$  through the time-distance equation. The estimated leak locations were stored in matrices  $l_{ij}^{APA}$  and  $l_{ij}^{SPA}$  where each leak location was relative to the corresponding source sensor  $i$ .

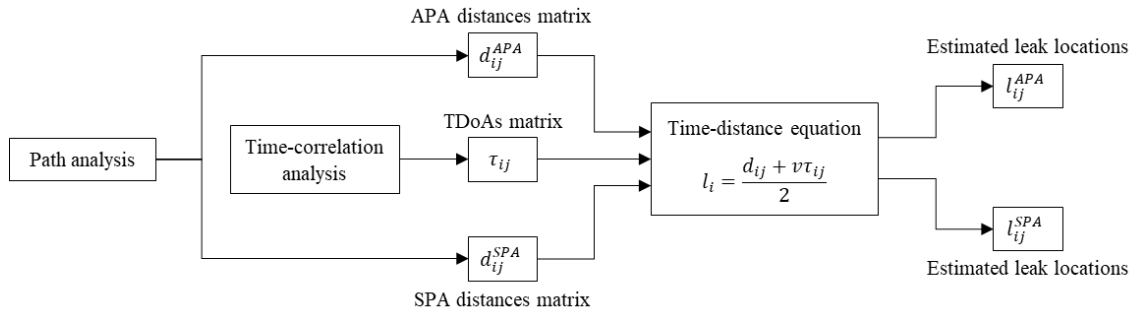


Figure 3.3: Location analysis based on APA and SPA.

However, the leak locations  $l_{ij}^{APA}$  and  $l_{ij}^{SPA}$  were inconclusive and hard to be compared statistically, as each leak location was only known by distance relative to respective source sensor. Thus, in the following section, the leak locations were generalised to a common source of reference through the backward path analysis and data reorganisation.

### 3.1.4 Backward path analysis and data reorganisation

Before the statistical analyses and pruning, backward path analysis and data reorganisation were introduced as pre-processing stages to generalised and redistribute the leak locations into classes of the common source of reference. Figure 3.4 shows the outline of backward path analysis, where an example case of estimated leak locations correlated from sensors 2 and 4 is used for demonstration.

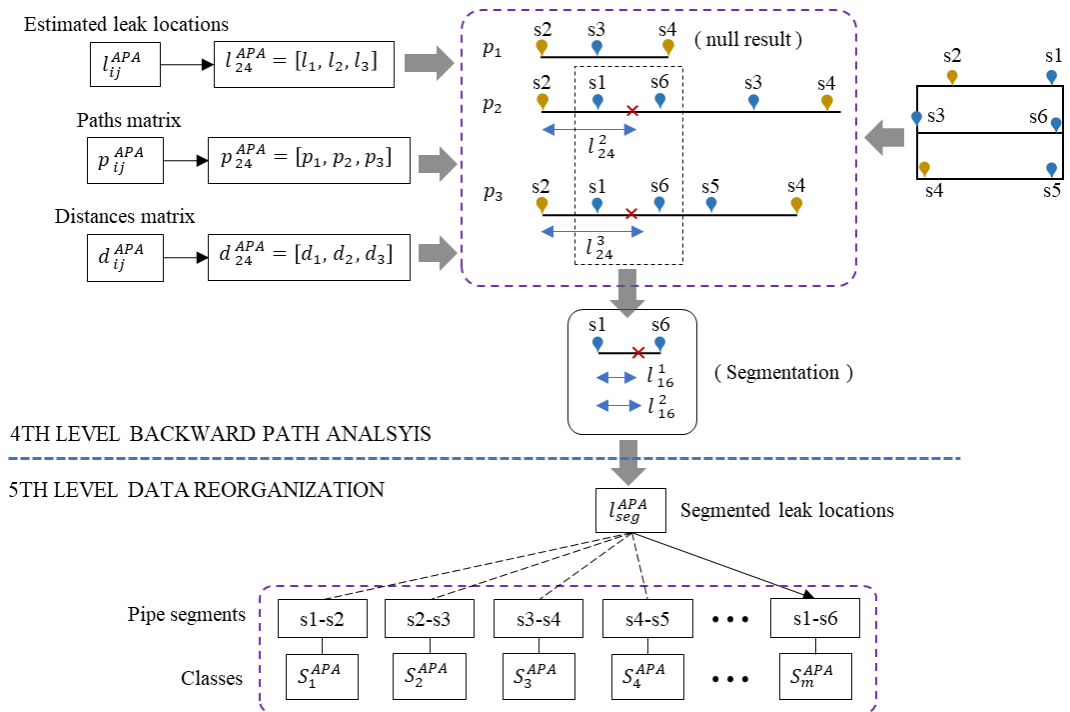


Figure 3.4: The backward path analysis and data reorganisation (example case: APA estimation results of correlation between sensors 2 and 4).

Each of the leak location was backtracked according to respective paths  $p_{ij}^{APA}$  and  $p_{ij}^{SPA}$  through the backward path analysis. The nearest pipe segment of the leak location was pruned from respective path through process of segmentation. The segmented leak locations,  $l_{seg}^{APA}$  and  $l_{seg}^{SPA}$  were represented by the distance from the nearest sensor in the pipe segment, and grouped by classes with respect to the pipe segment.

### 3.1.5 General statistical analysis

Figure 3.5 shows the reasoning processes of the MLAF, which comprises of three stages. General statistical analysis served as first screening to statistically deduce the final leak location from the estimations set. All the segmented leak locations were compared in terms of degrees of dominance and convergence at this stage. The degree of dominance,  $\Delta_{max}$  was determined by calculating the difference between the mode frequency,  $N_{mode}$  and the mean of all classes. The value tells us how much the modal class dominates the total estimations over the standard mean.

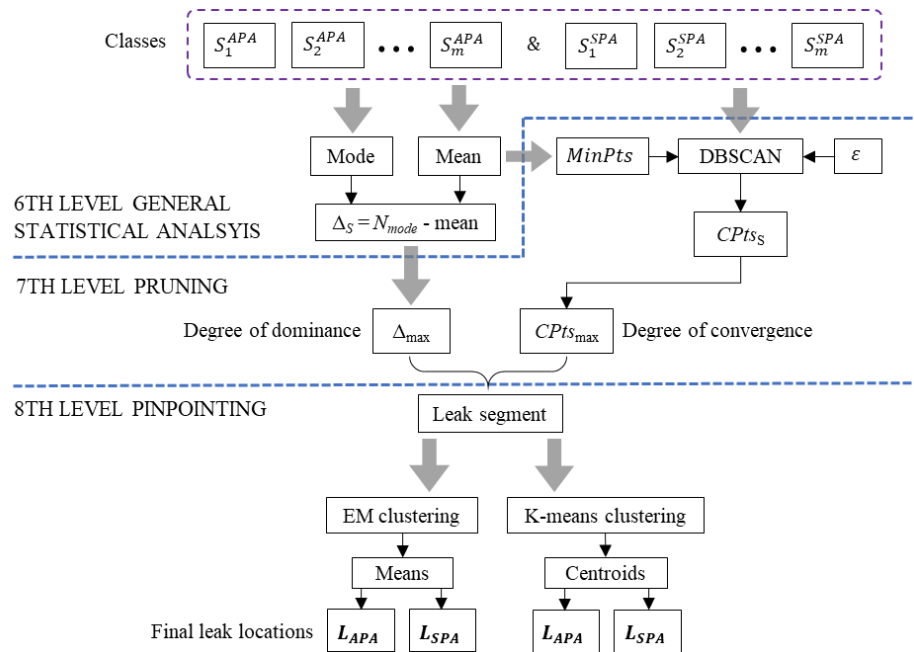


Figure 3.5: General statistical analysis, pruning and pinpointing.

### 3.1.6 Pruning

In the subsequent pruning process, density-based spatial clustering of applications with noise (DBSCAN) was implemented to further identify the density of class of segmented leak locations. Clusters were formed by grouping  $MinPts$  numbers of neighbouring data points within a maximum radius  $\varepsilon$ . The cluster with the highest density (or number of cluster points,  $CPts$ ) shows superior degree of convergence among the estimations set. Thus, at the end of the pruning process, the pipe segment with the highest cluster's density ( $CPts_{max}$ ) and degree of dominance ( $\Delta_{max}$ ) was concluded as the.

#### 3.1.6.1 DBSCAN

DBSCAN is a density-based clustering method to group a set of observations into one or more regions of high density based on two parameters: the radius,  $\varepsilon$  and  $MinPts$  (Liu & Özsu, 2009). Figure 3.6 depicts the DBSCAN result based on  $MinPts = 3$  and radius =  $\varepsilon$ , the cluster with number of points,  $CPts > MinPts$  forms a 'circle' centres at a core point, where the border points are the rest of the points lie within the radius =  $\varepsilon$ . The data points in low density region do not form a cluster and known as the noise points (outliers).

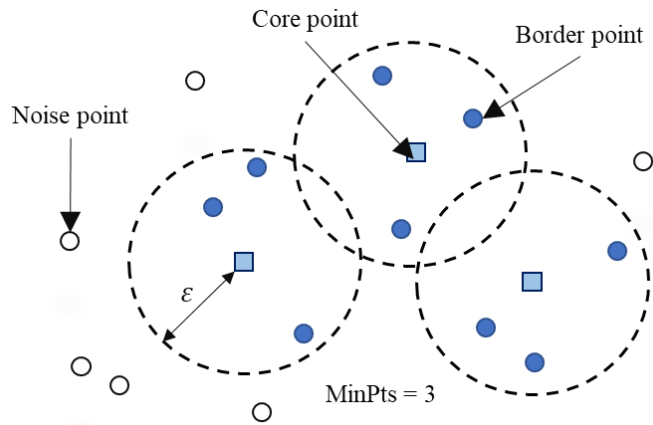


Figure 3.6: Core points, border points, and noise points (outliers).

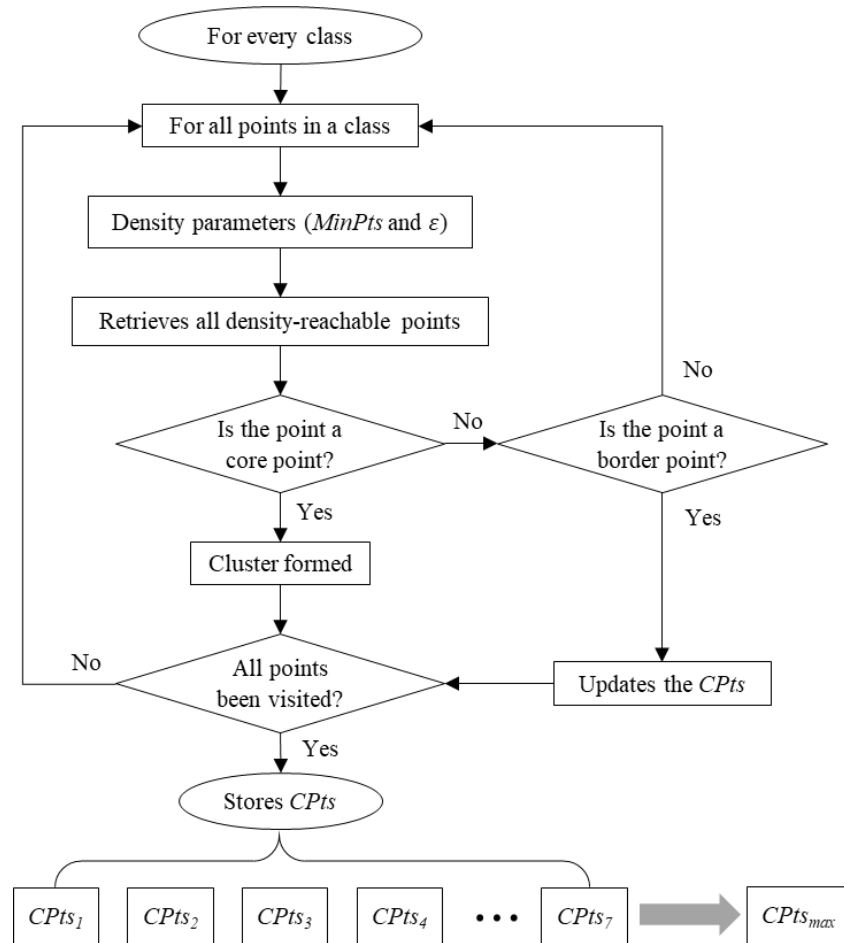


Figure 3.7: DBSCAN clustering process.

Figure 3.7 shows the process of the DBSCAN in determining the  $CPts$ . The process searched through all density-reachable points within a dataset and



determined whether the criteria  $CPts > MinPt$  based on the  $\varepsilon$  were met. All the  $CPts$  (core point and border points) of the cluster were stored and the process kept repeating until all points in a dataset in all available classes (pipe segments) were visited. The cluster with the highest density,  $CPts_{max}$  has the highest degree of convergence among the estimations set. At the end of this stage, the pipe segment with the highest degrees of dominance and convergence was deduced as the final leak segment.

### **3.1.7 Pinpointing**

In the next stage, the leak location was further pinpointed within the final leak segment based on expectation maximisation (EM) and K-means clustering. The segmented leak locations were grouped into clusters based on the centricity of the given distributed set. Each cluster was represented by the respective central vectors (EM mean and K-means centroid), where the final leak locations,  $L_{APA}$  and  $L_{SPA}$  were statistically concluded by these central vectors

#### **3.1.7.1 Expectation Maximisation Clustering**

EM clustering is an algorithm commonly used to partition  $i$  observations into cluster by performing maximum likelihood estimation in statistical models based on unobserved latent variables. The clustering process takes place by iteratively alternating the expectation (E) step and maximisation (M) step. In

the general case, with observations  $x_i$  and latent variables  $z_i$ , we have the log-likelihood function as follows:

$$l(\theta) = \sum_{i=1}^N \log p(x_i, z_i | \theta) \quad (6)$$

Where  $\theta = (\mu_i, \sigma_i^2, \pi)$ , means  $\mu_i$  and variances  $\sigma_i^2$  are the Gaussians parameters,  $\pi$  is the categorical variables. In E step, the log likelihood estimation function of  $\theta$ , with respect to the current conditional distribution of  $z_i$  given  $x_i$  and the current estimates of the parameters  $\theta_t$  is given by:

$$Q(\theta | \theta_t) = \sum_{i=1}^N E[\log[p(z_i | \theta) p(x_i, z_i | \theta)]] \quad (7)$$

In M step, the log likelihood estimation function is maximised as:

$$\theta_{t+1} = \arg \max Q(\theta | \theta_t) \quad (8)$$

The parameters in the maximised log likelihood estimation function are used to determine the distribution of the latent variables in the next E step, therefore maximise the log likelihood estimation function in E step. During each iteration, parameters  $\mu_i$  and  $\sigma_i^2$  are updated until  $\sigma_i^2$  have been minimised. The final leak location of each segment is represented by means  $\mu_i^t$ .

### 3.1.7.2 K-means clustering

K-means clustering is another unsupervised learning method to partition  $i$  observations into  $k$  clusters based on their nearest mean. The algorithm alternates between the calculation of the centroid and the sum of squares error (SSE). Initial search of centroid  $c_i$  starts with a random guess.

Equation (9) calculates the SSE between set of data points and centroids. During each iteration, new centroids of clusters are updated by equation as the mean values (10) until SSEs have been minimised. The final leak location of each segment is represented by a central vector (centroid) of each cluster,  $c_i$ .

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \|x - c_i\|^2 \quad (9)$$

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x \quad (10)$$

Where  $x$  is set of data points,  $C_i$  is  $i^{th}$  cluster,  $c_i$  is centroid of cluster  $C_i$ ,  $m_i$  is the number of points in  $i^{th}$  cluster,  $k$  is number of clusters.

### 3.2 MDNN

In this section, the MDNN was designed based on the inverse problem analysis methodology to localise a leak in piping network. The leak localisation was achieved by a series of custom training based on a few sets of training data. The training was conducted through supervised learning based on input features such as the TDoAs, receival sensors, and distances between sensors. The outline of the MDNN localisation is depicted in Figure 3.8. The overall architecture comprises of 3 levels: data preparation, MDNN training, and field prediction.

In 1<sup>st</sup> level data preparation, multiple sets of distributed leaks were emulated through the inverse location analysis. The leak locations were emulated with an equal separation distance along all pipe segments in the targeted piping network. Each leak location and the corresponding TDoA

served as the training data for an individual cases of multiple single-leak events at different locations.

In 2<sup>nd</sup> level MDNN's training, the MDNN was trained to identify the leak segments and locations through recognition of the input features. The training was repeated with different hyperparameters such as tensor shapes (numbers of hidden layers and dropout layers), number of epochs, batch sizes and the sample size of training data to identify the optimal model.

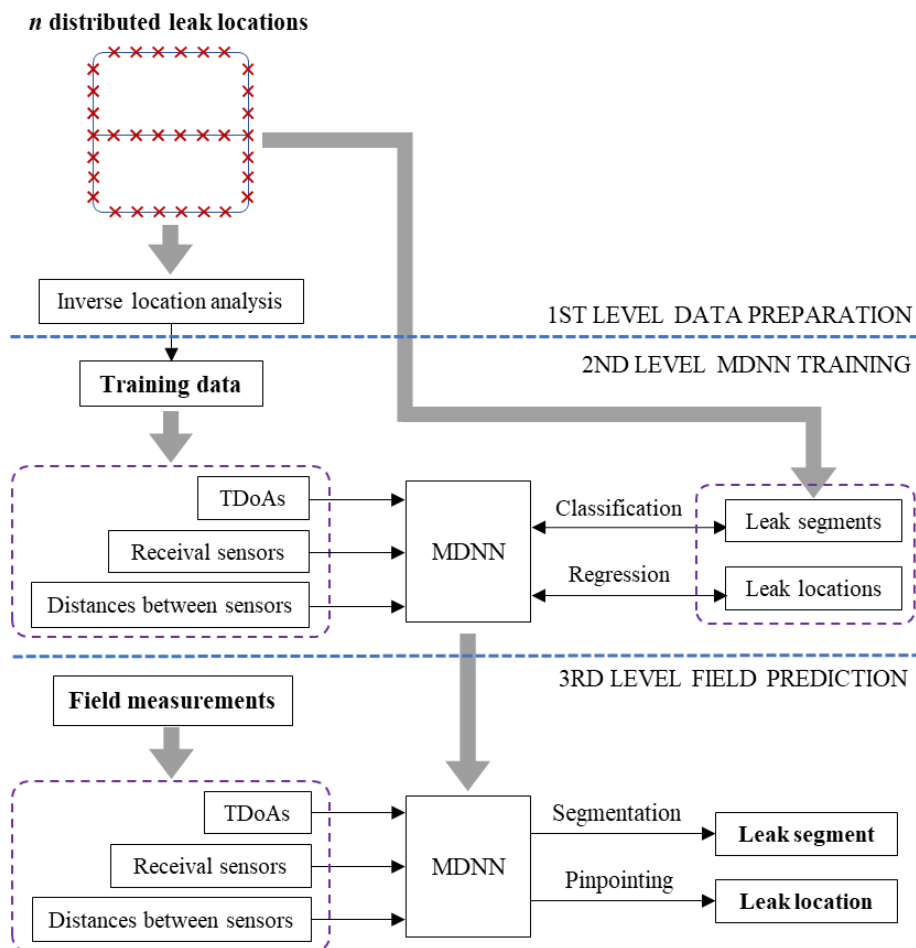


Figure 3.8: Outline of MDNN localisation.

Lastly, the optimal model was implemented for the field prediction in a local DMA. The prediction of actual leak was based on the field measured TDoAs as discussed in section 3.1.2, and the corresponding features (receival

sensors and the distances between sensors). The leak segment and location were predicted based on the optimal trained model.

### 3.2.1 Feed-forward and back-propagation

Feed-forward and back-propagation are the two essential operational mechanisms of the neural network training. These two mechanisms describe how the data spread through a network and improve the prediction gradually during the training. A basic deep neural network model consists of multi-layer feed-forward backpropagated ANNs with multiple hidden layers between the input and output layers. Figure 3.9 illustrates the basic deep neural network model with an input layer,  $n$  hidden layers and an output layer.

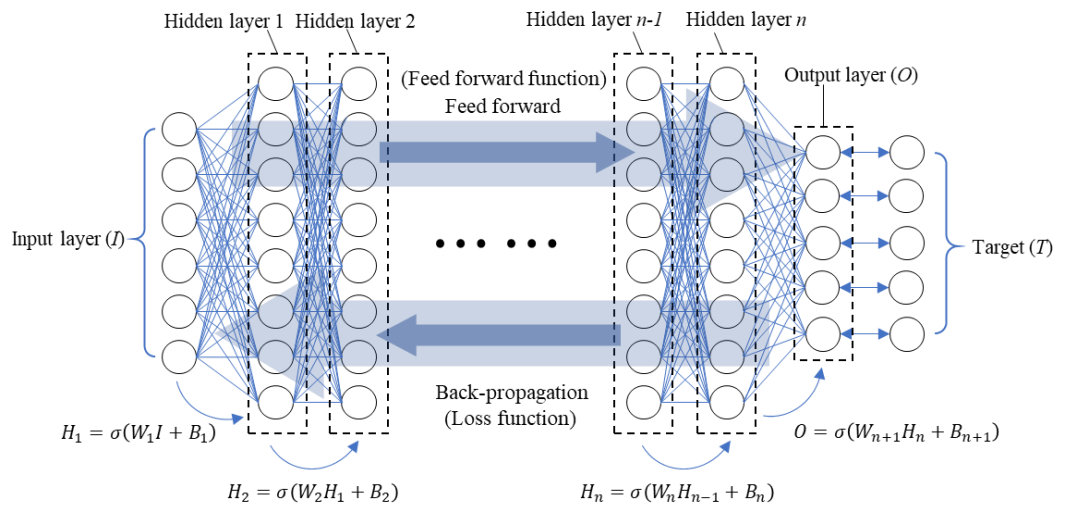


Figure 3.9: Basic neural network model.

Each neuron in the input layer holds a single value of input feature. The data is passed from the previous layer to the following layer by mean of the feed-forward mechanism. The general feed-forward function can be represented by equation (11), where  $x_{i-1}$  represents the neuron in previous layer,  $y_i$  is the

neuron in the subsequent layer,  $\sigma$  is the activation function,  $W_i$  is the weightage, and  $B_i$  is the bias.

$$y_i = \sigma(W_i x_{i-1} + B_i) \quad (11)$$

The values of output neurons vary with  $W_i$  and  $B_i$  and be normalised by the activation function,  $\sigma$  during the feed-forward process. At the end of the feed-forward process, quadratic error between the network output and respective targets,  $E$  are calculated through loss function. In the back-propagation process, the  $W_i$  and  $B_i$  are adjusted with the correcting factors  $\Delta W_i$  and  $\Delta B_i$ . The correcting factors  $\Delta W_i$  and  $\Delta B_i$  can determined from the partial derivative of the error function as in equations (12) and (13), respectively.

$$\Delta W_i = -\gamma \frac{\partial E}{\partial W_i}, \text{ for } i = 1, 2 \dots n \quad (12)$$

$$\Delta B_i = -\gamma \frac{\partial E}{\partial B_i}, \text{ for } i = 1, 2 \dots n \quad (13)$$

In the subsequent feed-forward process, the neural network continues to estimate the new outputs based on the adjusted activation function. The feed-forward and back-propagation mechanisms are repeated iteratively until the training is optimised. Equation (14) indicates the gradient  $\nabla E$  which is a continuous differentiable function of all weights  $W_1, W_2, \dots, W_n$  in the network. The training of neural network is optimised as the minimum of the error function,  $\nabla E = 0$  is achieved.

$$\nabla E = \left( \frac{\partial E}{\partial W_1}, \frac{\partial E}{\partial W_2}, \dots, \frac{\partial E}{\partial W_n} \right) \quad (14)$$

### 3.2.2 Mixed-model architecture

To effectively localised leak location in 2-dimensional layout piping network, the MDNN was constructed for simultaneous predictions of both pipe segment and location. The prediction involved both classification and regression tasks. Thus, two neural network architectures were aggregated to achieve the simultaneous predictions.

Classification neural network (CNN) was utilised as a classifier in predicting the leak segments, whereas regression neural network (RNN) was implemented to pinpoint the leak location. Figure 3.10 depicts the mixed-model architecture of MDNN which comprises four working subnetworks: the 1st level CNN, 1<sup>st</sup> level RNN, 2nd level CNN, and 2<sup>nd</sup> level RNN.

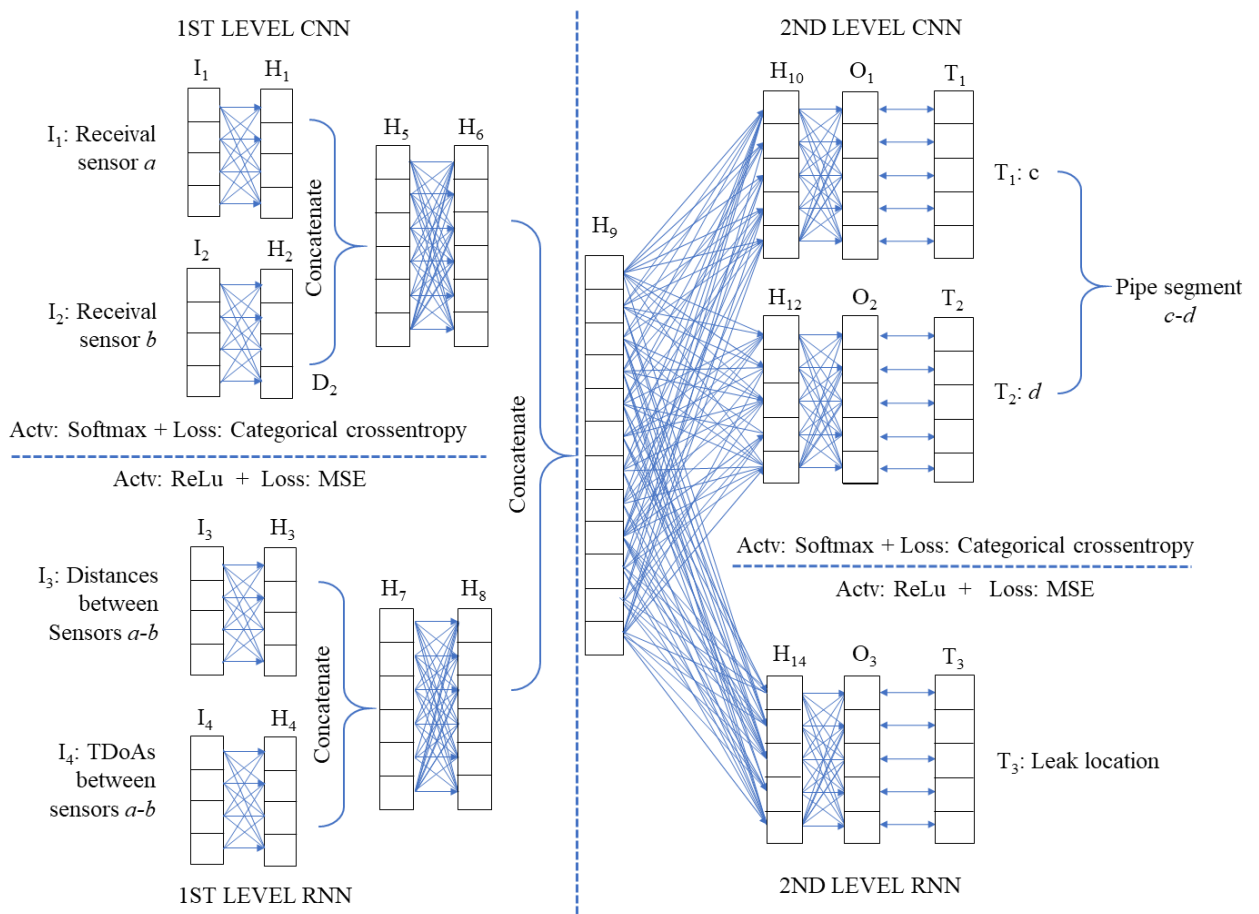


Figure 3.10: The mixed-model architecture of MDNN.

In 1st level subnetwork, the categorical input (receival sensors  $a$  and  $b$ ) and the numerical input (TDoAs and distances between each pair of receival sensors) were fed into the CNN and RNN, respectively. In 1st and 2<sup>nd</sup> level subnetworks, all the neurons of the previous layers were fully connected to subsequent layers. The 1st level subnetworks were linked to the 2<sup>nd</sup> level subnetwork by concatenating the output neurons in the 1<sup>st</sup> level subnetworks. The concatenation allows aggregated features learning based on the intermediate result from the preliminary learning of the four primitive input features. In 2<sup>nd</sup> level subnetworks, the leak segment and location were predicted by the CNN and RNN, respectively.

### 3.2.2.1 CNN

CNN is commonly utilised for prediction of categorical data. In MDNN, the receival sensors  $a$  and  $b$  in numerical labels  $(1, 2, 3, \dots, n)$  were one-hot encoded as categorical data and input into 1<sup>st</sup> level CNN. The feed forward and back-propagation processes were based on the Softmax activation function and categorical cross-entropy loss function, which are equated in eq. (15) and (16), respectively. Where  $S(x_i)$  is the Softmax function,  $m$  is the number of samples, and  $t_i$  is the actual probability of the targeted class. The Softmax function normalise the output of neuron to values of 0 and 1 so that the prediction is always converged. In the backpropagation process, the cross-entropy loss function measures the loss of respective output based on its probability values between 0 and 1.



$$S(x_i) = \frac{e^{x_i}}{\sum_i^m e^{x_i}} \quad (15)$$

$$L_{CE} = -\frac{1}{m} \sum_i^m t_i \log S(x_i) \quad (16)$$

### 3.2.2.2 RNN

The 1<sup>st</sup> and 2<sup>nd</sup> level RNNs are commonly governed by the rectified linear unit (ReLU) activation function and mean square error (MSE) loss function. In most of the cases, vanishing gradient problem is generally faced by RNN training, as the gradient of MSE loss function tends to become vanishingly small. The vanishing gradient problem inhibits the weights of a RNN from changing their values and leads to non-progressive training. ReLU function as expressed in eq. (17) is a piecewise linear function to exclude the negative output in the RNN. The function is well-known as an effective solution for the vanishing gradient problem with validated applications [23, 24]. Equation (18) further shows the MSE loss function utilised in the RNN, where  $m$  is the number of samples,  $t_i$  is the target value, and  $y_i$  is the predicted output of regression neural network.

$$y_i = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$$L_{MSE} = \frac{1}{m} \sum_i^m (t_i - y_i)^2 \quad (18)$$

### 3.2.3 Overfitting in neural network

One of the most common problems in the deep neural network training is overfitting. There are several reasons that a deep neural network can overfit. A noisy training dataset or highly complex model with too many parameters can eventually causes the deep neural network to overfit. Figure 3.11 shows an example of the overfitting issue in training which can solved by stopping the training at an optimal point. The solution is commonly known as the ‘early stopping’. In general, there are several ways of preventing overfitting in a deep neural network, including simplify the complexity of the neural network by changing tensor shape, introducing dropout layers, changing the training batch sizes and ‘early stop’.

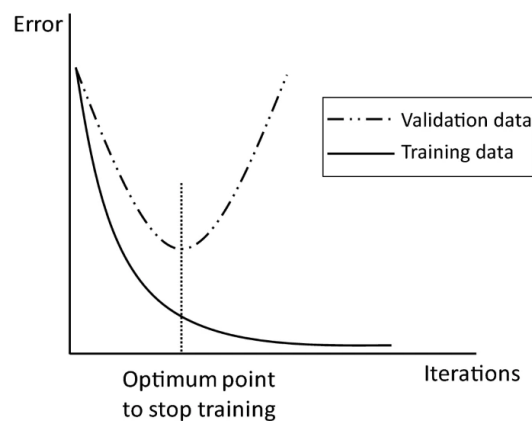


Figure 3.11: Overfitting and early stopping (Serveh, et al., 2019).

#### 3.2.3.1 Number of epochs

Epoch is a complete process of feed-forward and back-propagation in a neural network. Number of epochs refer to the number of iterations the neural network went through the process. The number of epochs is conventionally large, often hundreds or thousands, to ensure the sufficient training. As

discussed in previous section, overfitting is a very common issue in neural network training. Early stopping can be achieved by limiting the number of epochs in training.

### **3.2.3.2 Tensor shapes**

Modifications of the neural network's tensor shapes is another effective ways to prevent the neural network from overfitting. The neural network can overfit during a training due to the network complexity. Therefore, the key ideas are to reduce the network complexity by changing the number of hidden layers and the number of neurons in each layer to obtain the optimal tensor shape for training.

### **3.2.3.3 Dropout**

Dropout is a regularisation technique to prevent the neural networks from overfitting. The technique modifies the network layout by randomly dropping some connections of the neurons from each layer during the training. The technique is equivalent to the modification of tensor shapes with a constant number of layers, as the training is performed in different 'dropped' neural networks in every epoch. Figure 3.12 shows an example of dropout during a neural network training. This technique has proven to degrade the overfitting issue in a variety of problems such as image segmentation, image classification, word and semantic embedding.

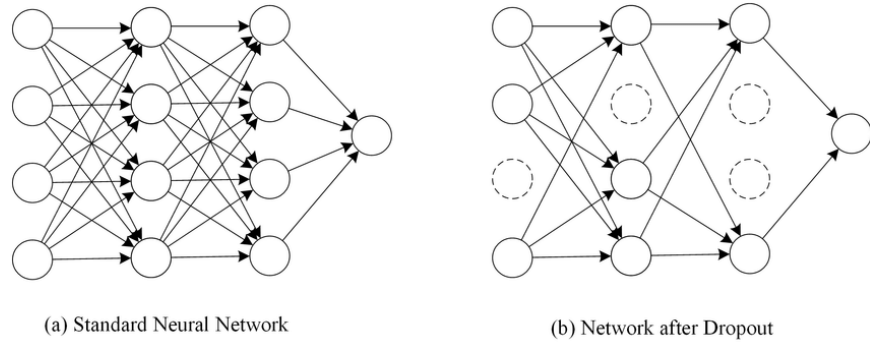


Figure 3.12: ropout of neural network (Mohamed & Hichem, 2018)

### 3.3 Mathematical modelling of the arrival times

In this section, a mathematical modelling technique was proposed to simulate leak data for the characterisation tests of MLAF and the MDNN training. It is an inevitable challenge to characterize and validate the performance of MDNN, due to limited access of piping networks with various topologies. The accesses to piping networks of varying network sizes, network shapes, and locations of leak are practically unachievable. Moreover, MDNN's training requires data of different leak locations in the targeted piping network, which is simply impossible to be achieved experimentally. Therefore, the mathematical modelling was proposed as an alternative to mimic the real measurement of leak in a virtual piping network and simulate the temporal data of the arrival time for the characterisation tests of MLAF and training data in MDNN.

Figure 3.13 shows a sample case of leak data simulation with consideration of the multi-directional waves issue. During leakage, the multi-directional waves,  $x_1, x_2, x_3, \dots, x_n$  from the leak source tend to arrive at two individual sensors 1 and 2 at different times, due to different in transmission

paths. The arrival times of multi-directional waves were therefore noted as  $t_1$ ,  $t_2$ ,  $t_3$ ,  $\dots$ ,  $t_n$ , respectively. Based on the superposition theorem of waves, the resultant leak wave is expected to arrive at sensors 1 and 2 in the resultant waves  $x_{s1}$  and  $x_{s2}$ . The waves representation of  $x_{s1}$  can thus be derived from the superposition of waves,  $x_1$ ,  $x_2$ ,  $x_3$ ,  $\dots$ ,  $x_n$  as in eq. (19). Each of the wave function carry the respective phases shifts,  $\alpha_2$ ,  $\alpha_3$ ,  $\dots$ ,  $\alpha_n$  resultant from the time differences, due to the respective differences in path lengths.

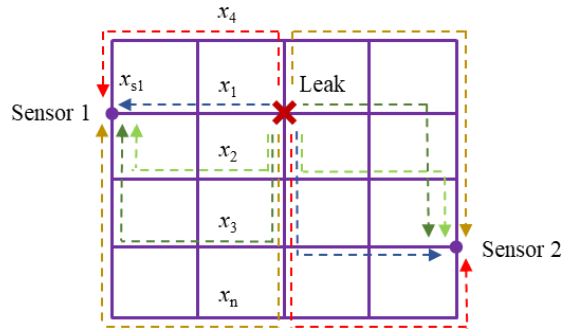


Figure 3.13: Sample case of leak data simulation.

$$x_{s1}(t) = \begin{cases} x_1(t_1) = x_0 \sin(\omega t_1) \\ x_2(t_2) = x_0 \sin(\omega t_1 + \alpha_2) \\ \vdots \\ x_n(t_n) = x_0 \sin(\omega t_1 + \alpha_n) \end{cases} \quad (19)$$

$$\alpha_2 = 2\pi f \Delta t_2, \quad \Delta t_2 = t_2 - t_1 \quad (20)$$

$$\alpha_n = 2\pi f \Delta t_n, \quad \Delta t_n = t_n - t_1 \quad (21)$$

$$x_{s1}(t) = x_1(t_1) + x_2(t_2) \dots + x_n(t_3) \quad (22)$$

The resultant wave at sensor 1,  $x_{s1}$  and the phases shift,  $\alpha$  can thus be written as in eq. (23) and (24). The resultant wave at sensor 2,  $x_{s2}$  can be derived through the similar method. Finally, the TDoAs between resultant waves  $x_{s1}$  and  $x_{s2}$  can be calculated by subtracting respective arrival times  $t_{s1}$  and  $t_{s2}$  as in eq. (26).

$$x_{s1}(t_{s1}) = x_0(t) \sin(\omega t_1 + \alpha) \quad (23)$$

$$\alpha = \tan^{-1} \left[ \frac{\sum_{i=2}^n \sin \alpha_i}{\sum_{i=2}^n \cos \alpha_i} \right], \quad \alpha = 2\pi f \Delta t_{s1} \quad (24)$$

$$\Delta t_{s1} = t_{s1} - t_1, \quad t_{s1} = t_1 + \Delta t_{s1} \quad (25)$$

$$\tau_{12} = t_{s1} - t_{s2} \quad (26)$$

In part 3.13, the acoustic signals of leak were retrieved experimentally and the TDoA between two acoustic signals was determined through a recursive time-correlation analysis. In this part, the acoustic signals and TDoA were simulated based on the proposed mathematical modelling. The mathematical modelling provides an alternative solution for the characterisation tests and training data which are impractical and time-costly to be achieved experimentally.

### 3.4 Summary

This this chapter, two leak localisation methodologies based on the forward and inverse analysis methods were proposed. The detailed localisation processes of MLAF and MDNN have been discussed deliberately in Sections 3.1 and 3.2, respectively. The summary of the methodologies was depicted as flow diagram in Figure 3.14.

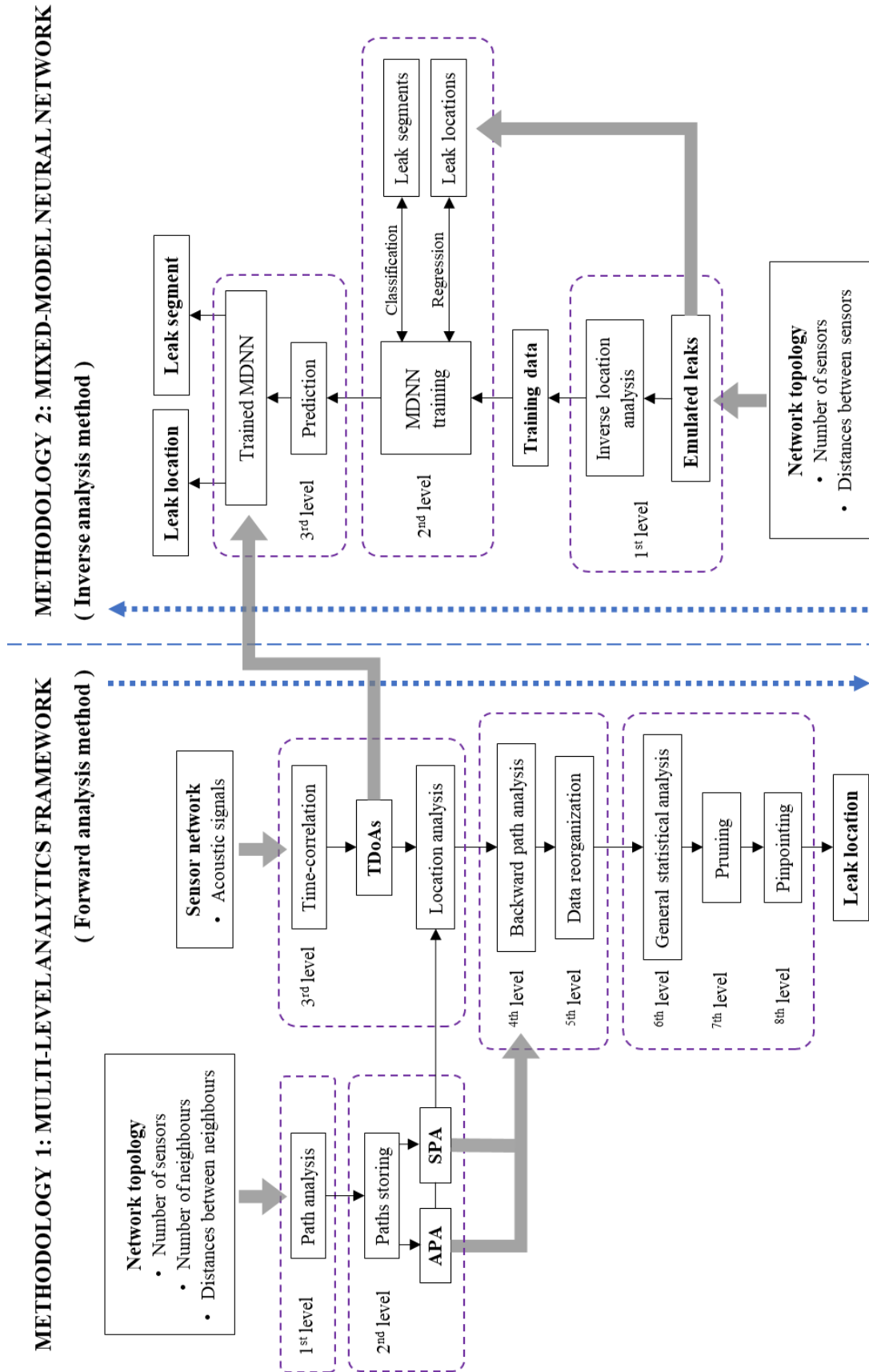


Figure 3.14: Summary of methodologies 1 and 2.

## **CHAPTER 4**

### **RESULTS AND DISCUSSIONS**

In the first part of this chapter, various characterisation tests were devised based on different topologies (sizes and shapes) of the piping networks to validate the performance and flexibility of the MLAF in different scenarios. The discussion was followed by a field prediction in a local DMA to further verify the reliability.

In the second part, the MDNN was trained to predict the leak location in similar DMA. The optimal model was applied for similar field prediction as in the first part. Finally, the localisation results of both methodologies were compared in terms of the localisation accuracy, processing time, and flexibility.

#### **4.1 MLAF leak localisation**

##### **4.1.1 Characterisation based on network sizes**

The first characterisation was conducted based on the network sizes, deviation of leak locations within each size of network, and the number of multi-



directional waves involved in leakage. 20 piping networks with different network sizes and leak locations were prepared for the characterisation tests.

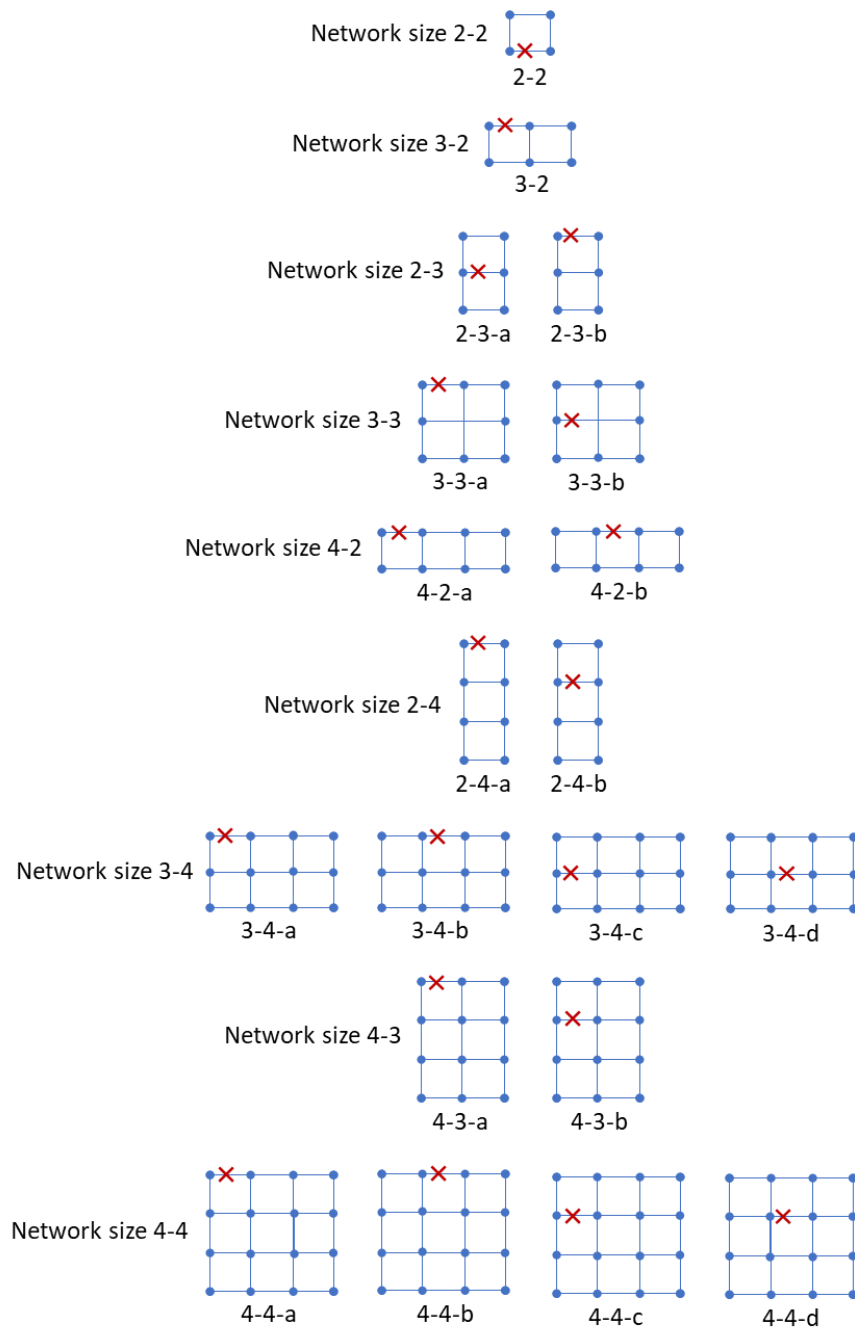


Figure 4.1: The 20 emulated piping networks with different network sizes and leak locations.

Figure 4.1 shows the 20 piping networks with varying network sizes from 4 sensors to 16 sensors. A consistent segment length of 50m was fixed in each network to study the sole effect of network sizes. Different positions of

leak were included in each size of the network to verify the possible effect to the localisation accuracy. Besides, a maximum number of 5 multi-directional waves was considered in each case of emulation to study the corresponding effects. The leak locations of all emulated cases were predicted using the MLAF.

The performances of MLAF were evaluated in two stages: first based on the degrees of dominance and convergence to evaluate the accuracy of pruning and followed by the RMSE to evaluate the accuracy of pinpointing. Figures 4.2 3 (a) and (b) show the degree of dominance and convergence show the first stage APA results. Both degrees of dominance and convergence of the APA results show increasing trend with higher number of multi-directional waves. However, both results do not show an obvious trend corresponding to varying network sizes and leak locations.

Figures 4.3 (a) and (b) show the degree of dominance and convergence of the first stage SPA results. The degree of dominance shows a decreasing trend while the degree of convergence remains constant. It is worth to note that the APA outperformed SPA in the pruning results, but the pruning based on SPA was still decent in all the tested cases. Besides, pruning results of both APA and SPA are not affected significantly by the size of network as there are no noticeable trend can be observed in Figures 4.2 (b) and 4.3 (b).

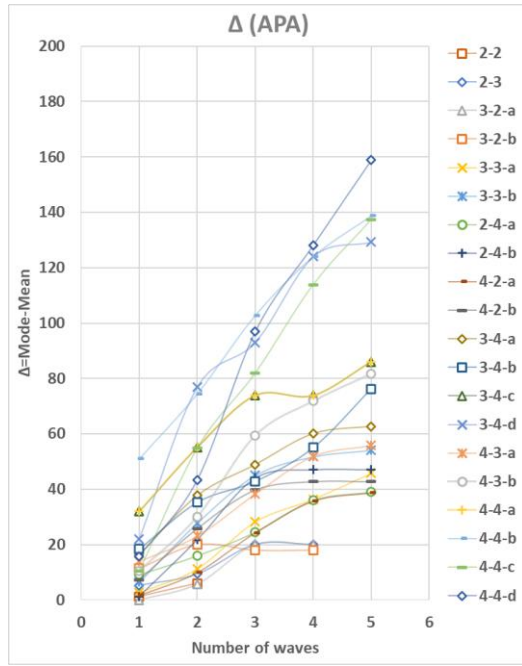


Figure 4.2 (a): First stage evaluations of APA results based on the degree of dominance.

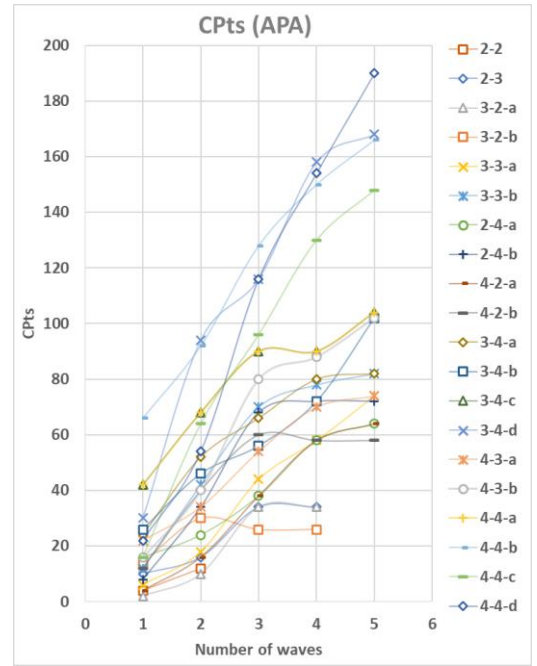


Figure 4.2 (b): First stage evaluations of APA results based on the degree of convergence.

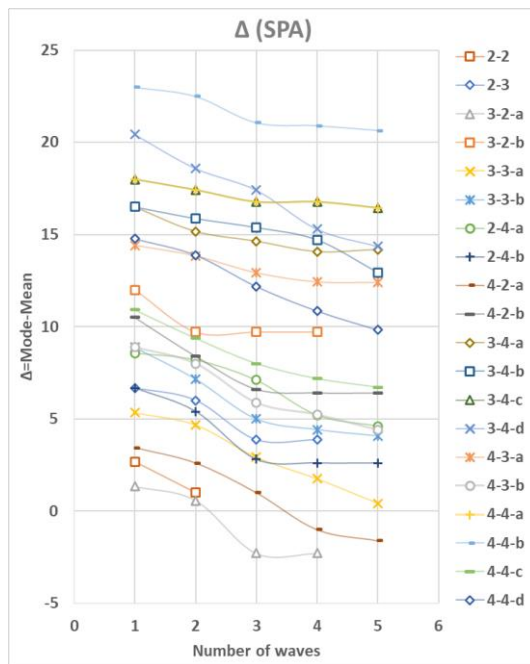


Figure 4.3 (a): First stage evaluations of SPA results based on the degree of dominance.

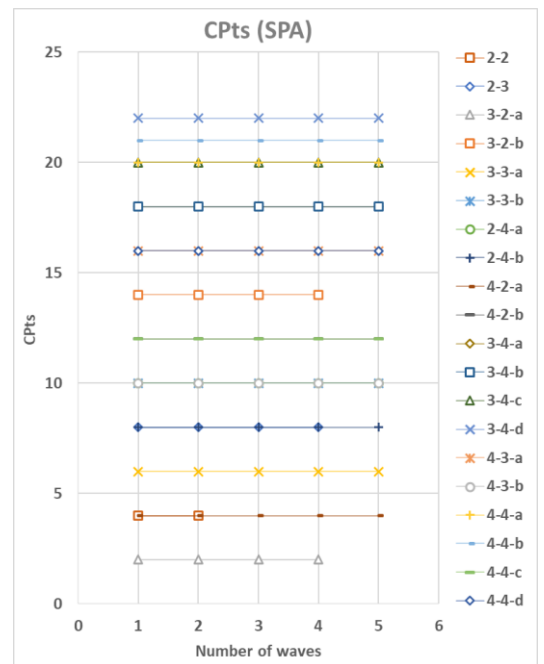


Figure 4.3 (b): First stage evaluations of SPA results based on the degree of convergence.

Subsequently, the results of localisation were further compared in term the RMSEs. The RMSEs of the final leak locations based on APA and SPA

were plotted in Figures 4.4 (a) and (b). The results show the localising accuracies of both APA and SPA did not vary significantly with different sizes of network and deviation of the leak locations. However, both APA and SPA had lower localisation accuracies when higher number of multi-directional waves involved in the leakage. A maximum RMSE  $> 0.9\text{m}$  were achieved by both APA and SPA localisations, which was considerably accurate in a 50 m pipe segment.

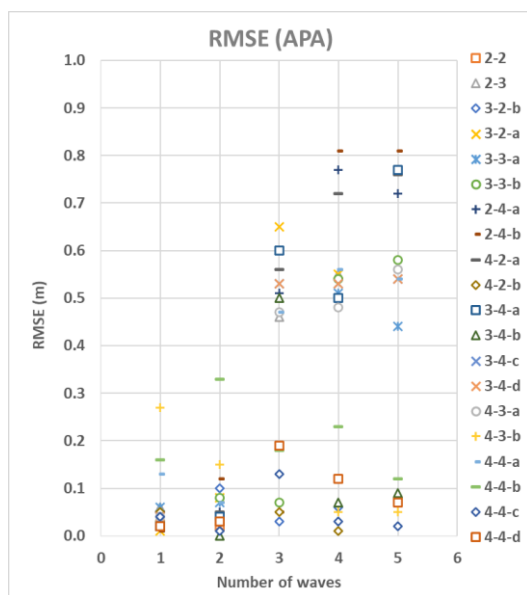


Figure 4.4 (a): Second stage evaluations of APA based on the RMSEs

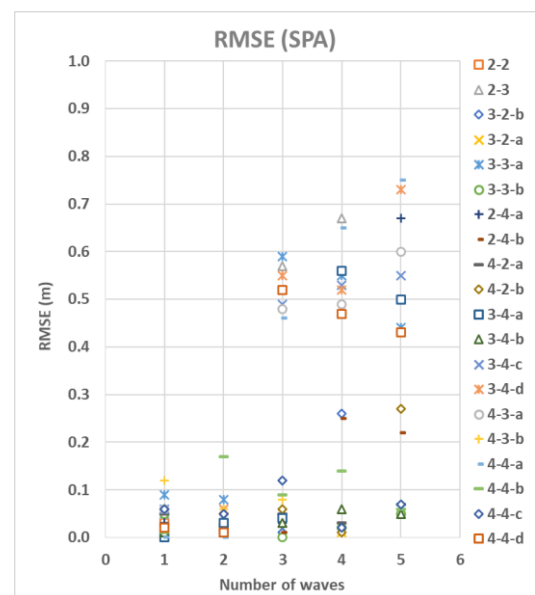


Figure 4.4 (b): Second stage evaluations of SPA based on the RMSEs

#### 4.1.2 Characterisation based on network shapes

The characterisation tests in the previous section have shown that the performances of MLAF is independent to the sizes of network and the different positions of the leak locations. The following characterisation was conducted to further study the performances against different network shapes. The characterisation tests also studied the effect when one side of the network is

deviating from another. Figure 4.5 shows five quadrilateral piping networks with a fixed size of 9 sensors. where each of the segments  $x$  and  $y$  were set with lengths  $y = 50$ ,  $x = ny$ , and the ratio  $n = 1, 2, 3, 4,$  and  $5$ . The extreme case such that  $x = 5y$  have been included in the characterisation tests. Four different leak locations  $a, b, c,$  and  $d$  were simulated in each network to characterise the localising accuracy when leak locations were strayed to one side of network.

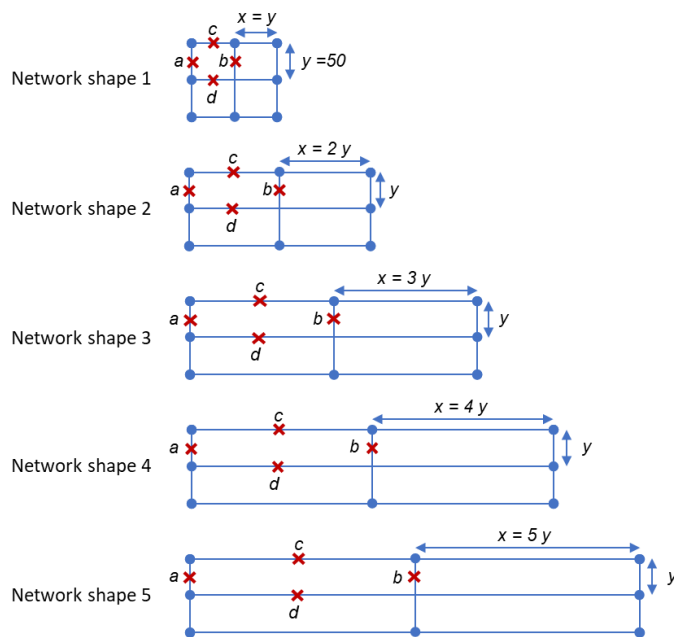


Figure 4.5: Five quadrilateral DMAs with different length ratios.

Figures 4.6 and 4.7 present the results of first stage evaluation based on the APA and SPA, respectively. The pruning results show similar trends with the previous characterisation tests corresponding to the number of multi-directional waves. However, the results did not vary significantly with different shapes of network and deviation of the leak locations, even in the extreme case of  $x = 5y$ . Nevertheless, the characterisation further affirmed the outperformance of the APA over the SPA in terms of the pruning.

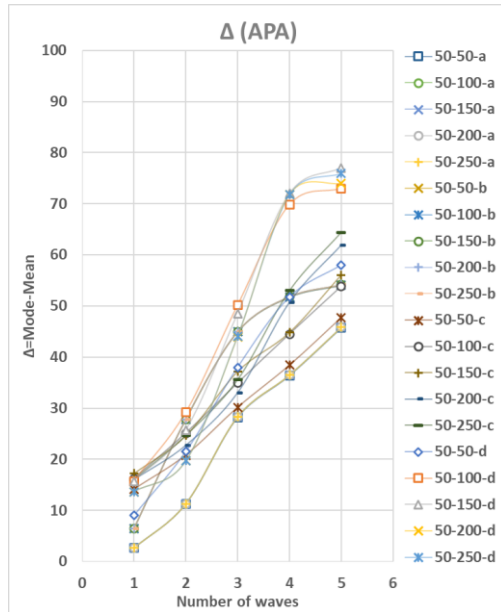


Figure 4.6 (a): First stage evaluations of APA results based on the degree of dominance.

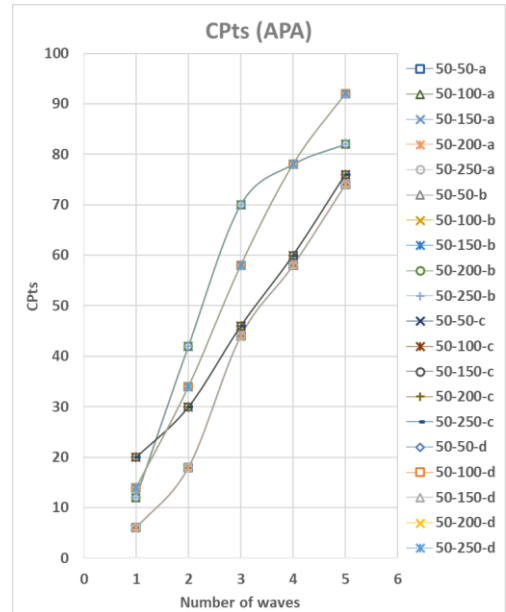


Figure 4.6 (b): First stage evaluations of APA results based on the degree of convergence.

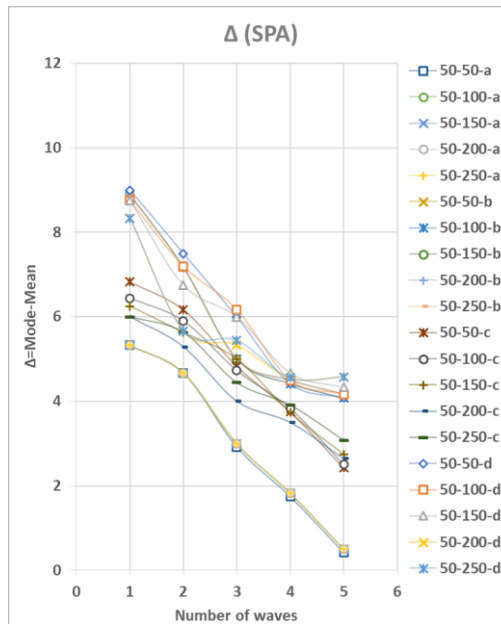


Figure 4.7 (a): First stage evaluations of SPA results based on the degree of dominance.

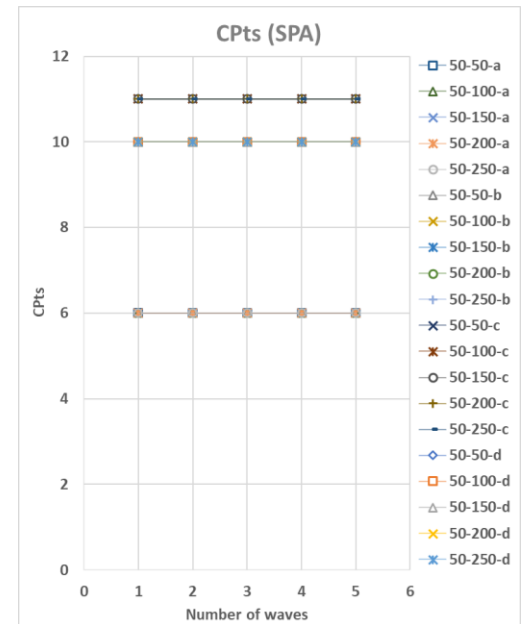


Figure 4.7 (b): First stage evaluations of SPA results based on the degree of convergence.

The RMSEs of leak locations  $a$ ,  $b$ ,  $c$ , and  $d$  are presented in Figures 4.8, (a), (b), (c) and (d), respectively. In general, the RMSEs show an increasing trend with higher number of multi-directional waves. Besides, a noticeable

deviation of leak locations  $c$  can be observed in Figure 4.8 (c). The RMSEs of leak locations  $c$  were comparatively large in network shapes 50-150, 50-200, and 50-250 when five multi-directional waves involved. However, the deviations in range of ( $1.5\text{m} < \text{RMSE} < 3\text{m}$ ) were insignificant in the pipe segment of length  $>150\text{m}$ , as the error were only  $>2\%$  of the total length.

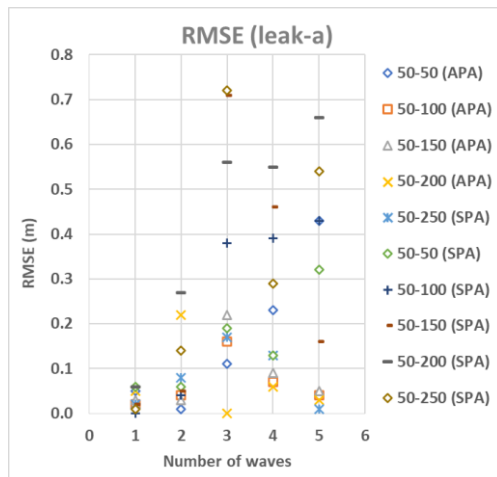


Figure 4.8 (a): Second stage evaluations based on the RMSEs of estimated leak location  $a$

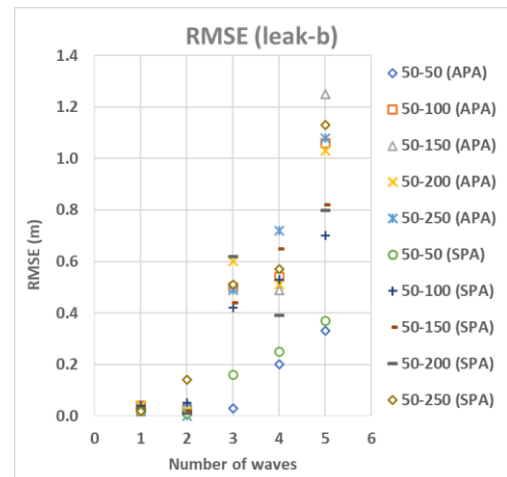


Figure 4.8 (b): Second stage evaluations based on the RMSEs of estimated leak location  $b$

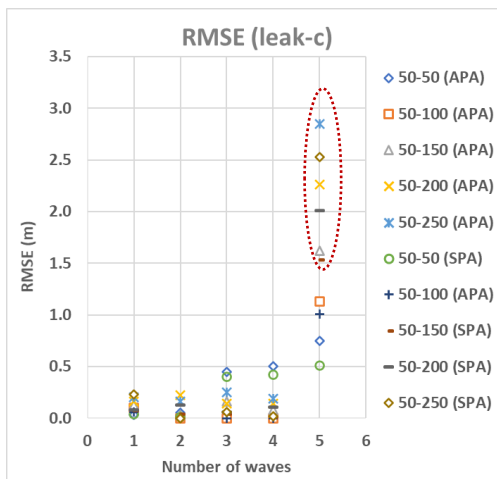


Figure 4.8 (c): Second stage evaluations based on the RMSEs of estimated leak location  $c$

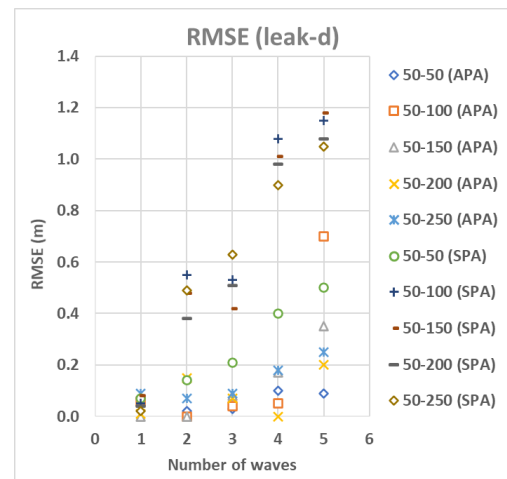


Figure 4.8 (d): Second stage evaluations based on the RMSEs of estimated leak location  $d$

### 4.1.3 Validation tests based on non-uniform DMAs

Excellent performances and the high adaptiveness of MLAF in various networks with different topologies have been proven by the previous characterisations. The results showed that APA and SPA were preferable in pruning and location pinpointing, respectively. Continuingly, two piping networks with non-uniform shapes and sizes were simulated to further validate the performances of the MLAF. Figure 4.9 depicts an emulated piping network with a single leak in segment 4-5. The leak location was set at 15m from sensor 4 and the case of leak was simulated based on three multi-directional waves.

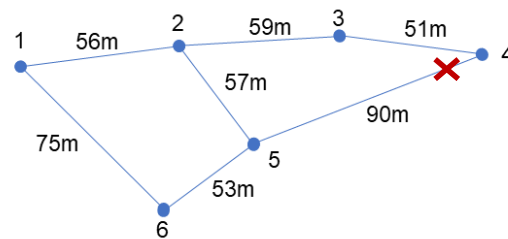


Figure 4.9: Emulated DMA 1.

The localisation results of general statistical analysis, pruning and pinpointing were plotted in Figures 4.10, 4.11, and 4.12, respectively. Significant results of pruning were achieved by the APA, with superior 30.3 degree of dominance and 34 degree of convergence in the leak segment 4-5. Figure 4.12 further shows the pinpointing result within the leak segment 4-5. The final leak locations were represented by the centroids of K-means and means of EM. The results were excellent with 0.38m and 0.31m RMSEs achieved by APA and SPA, respectively. In the first validation test, the MLAF localisation was time-effective with total processing time of 4.16s.



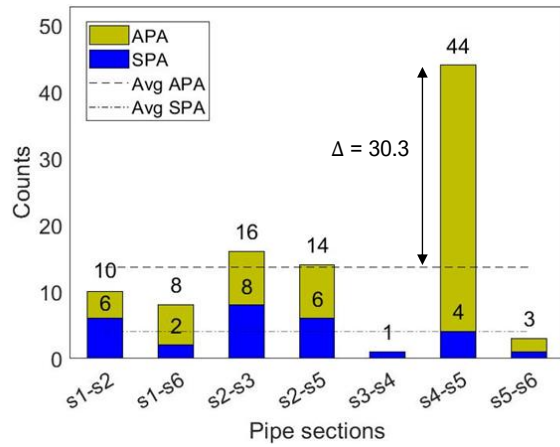


Figure 4.10: 6th level general statistical analysis.

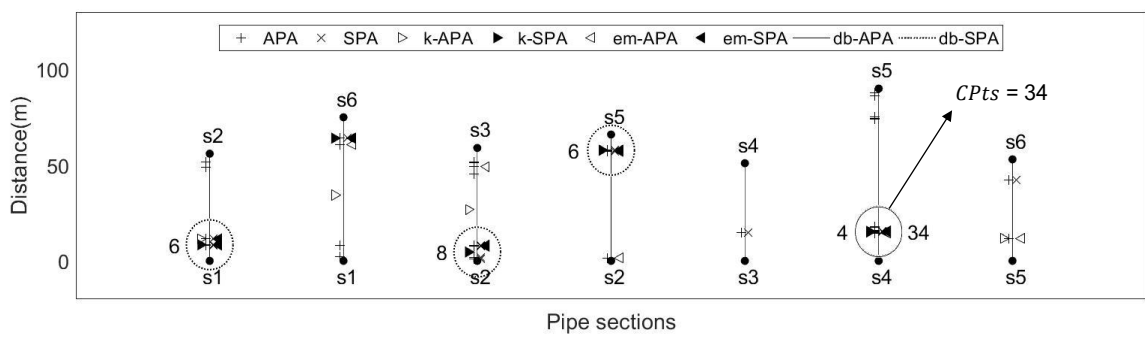


Figure 4.11: 7th level pruning.

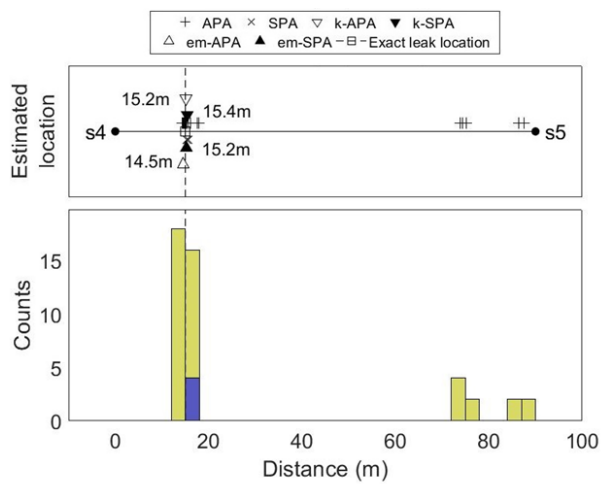


Figure 4.12: 8th level pinpointing.

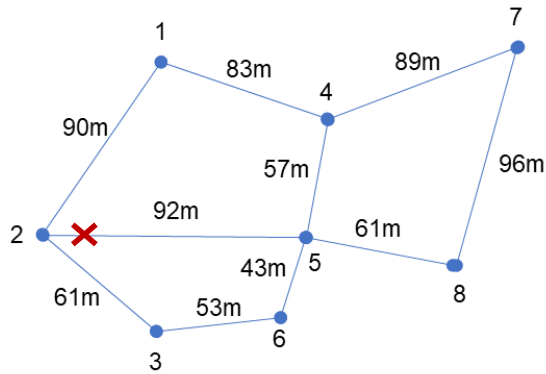


Figure 4.13: Simulated DMA 2 with non-uniform shape.

The validation test 2 was conducted based on emulated piping network 2 in Figure 4.13. A single leak location was simulated in segment 2-5, 15m from sensor 2, based on four multi-directional waves. The localisation results of general statistical analysis, pruning and pinpointing are presented in Figures 4.14, 4.15, and 4.16, respectively. The APA pruning was excellent with appreciable degree of dominance of 42.7 and degree of convergence of 64. The pinpointing results based on APA and SPA were both remarkable with RMSEs of 1.41m and 0.16m, respectively. The 4.74s localisation process was time-effective. The localisation results of validation tests 1 and 2 were summarised in Table 4.1.

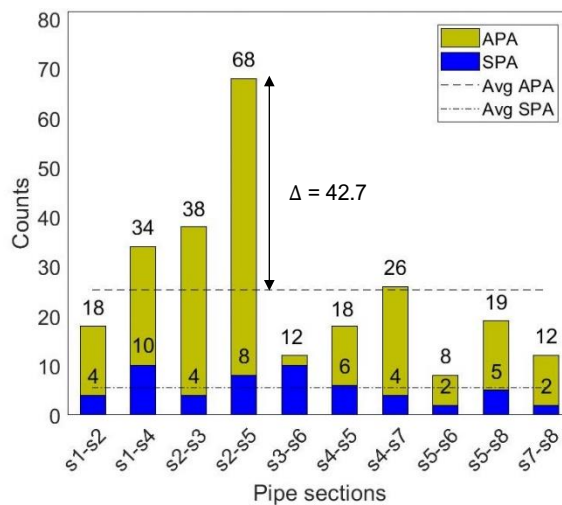


Figure 4.14: 6th level general statistical analysis.

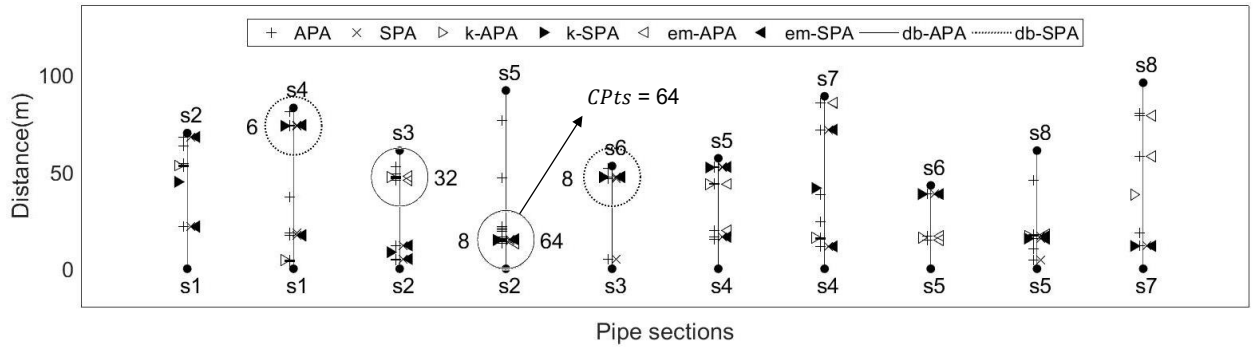


Figure 4.15: 7th level pruning.

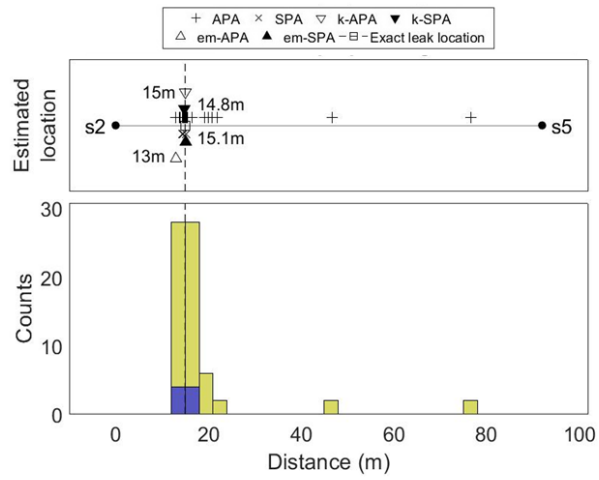


Figure 4.16: 8th level pinpointing.

Table 4.1: Localisation results of validation tests 1 and 2.

		Test 1	Test 2
$\Delta_{max}$	APA	30.3	42.7
	SPA	4	2.5
$Cpts_{max}$	APA	34	64
	SPA	4	8
K-mean (m)	APA	15.2	15.0
	SPA	15.4	14.8
EM-mean (m)	APA	14.5	13.0
	SPA	15.2	15.1
RMSE (m)	APA	0.38	1.41
	SPA	0.31	0.16
Time (s)		4.16	4.72

#### 4.1.4 Field prediction in the local DMA

A case study was conducted by experimenting the MLAF in a local DMA in Taman Scientex, Masai, Pasir Gudang, Johor (GPS coordinate:  $1^{\circ}30'42.3''\text{N}$   $103^{\circ}54'57.8''\text{E}$ ). A remote-acoustic sensor network consists of six hydrophone sensors was deployed in the DMA. The placement of the six sensors was shown in Figure 4.17. During the field test, each sensor was installed onto the underground valve of pipeline as shown in Figure 4.18. The acoustic signals of leak were collected through a 60s time-frame continuous data acquisition. Figure 4.19 shows the acoustic signal captured by six hydrophone sensors during the field test. Noticeable higher amplitudes can be observed in the signals from sensors s1, s6 and s2, as those sensors were closer to the leak source. Figure 4.20 further shows the time-correlation result of the leak signals from each pair of sensor in the sensor network.

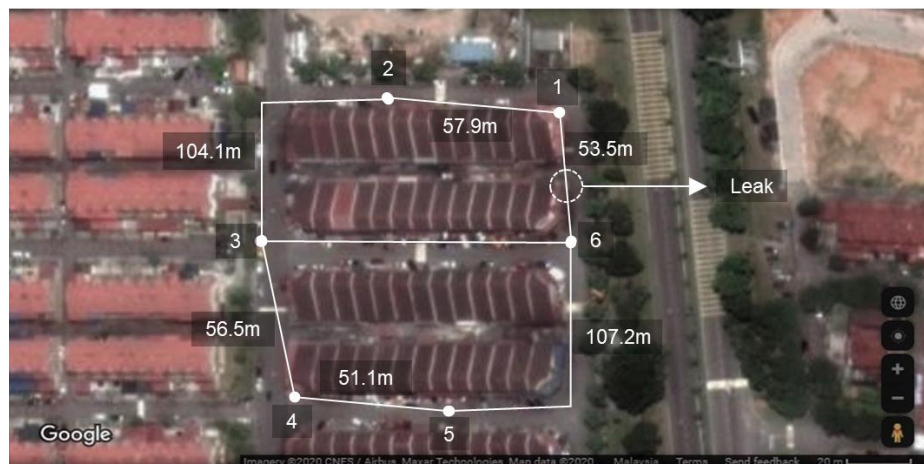


Figure 4.17: Local DMA at Jalan Pelanduk, Taman Scientex, Masai, Pasir Gudang, Johor. (GPS coordinate:  $1^{\circ}30'42.3''\text{N}$   $103^{\circ}54'57.8''\text{E}$ ).

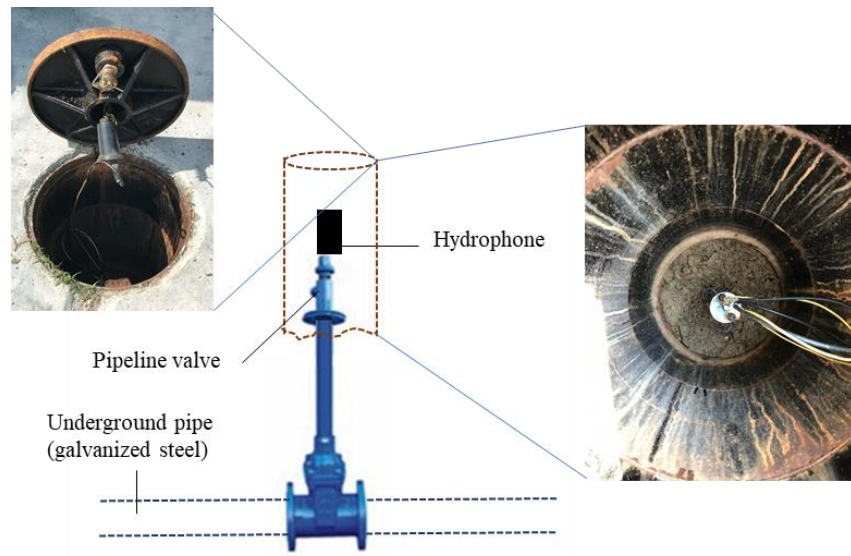


Figure 4.18: Installation of the hydrophone sensor.

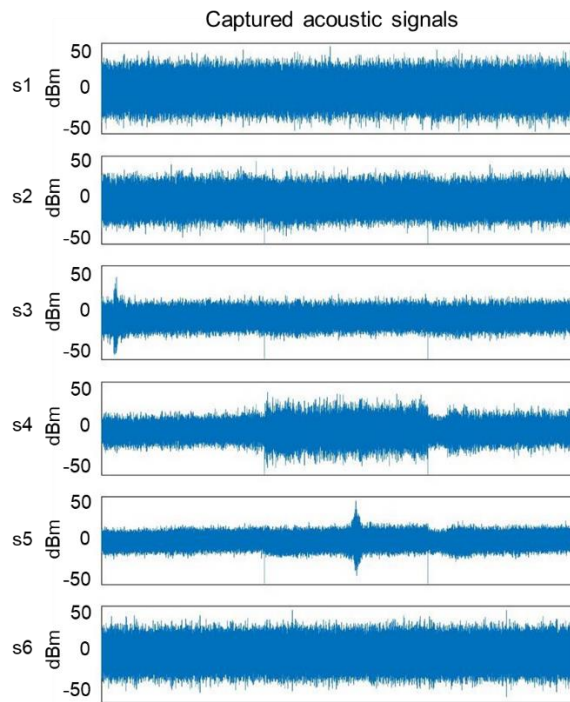


Figure 4.19: Acoustic signals captured by the six hydrophone sensors.

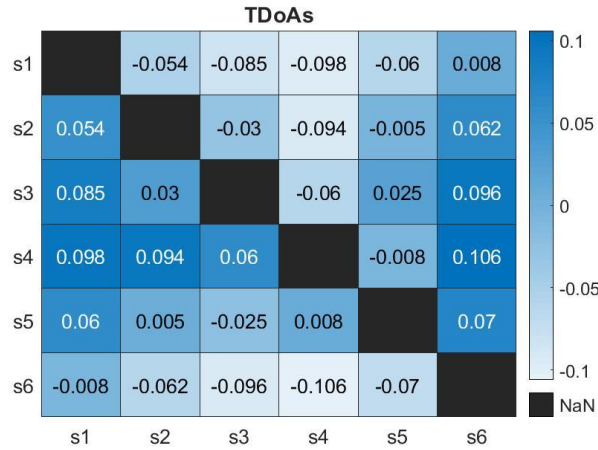


Figure 4.20: TDoAs between respective acoustic signals.

Figures 4.21 (a), 4.22 (a) and 4.23 (a) show the localisation results based on the field measurement. The results were compared to the localisation results based on the simulated input in Figures 4.21 (b), 4.22 (b) and 4.23 (b). The simulated leak input was prepared based on the mathematic modelling of three multi-directional waves.

Figures 4.21 (a) and (b) present the general statistical analysis based on the field measurement and simulated input. Both results had a close value of means and an identical mode in pipe segment s1-s6. The degrees of dominance of the simulated input and field measurement were considerably high, which were 10 and 20.1, respectively. Both of the results generally deducted the leak segment in s1-s6. The APA pruning based on the field measurement shows a minor confusion in pipe segment s3-s6 with a degree of dominance of 6. However, the results of SPA pruning are equivalent for both the field measurement and simulated input.

Figures 4.22 (a) and (b) further shows the pruning results based on degrees of convergence. The *CPTs* based on field measurement and simulate

input were 18 and 34, respectively. The high degrees of convergence clearly deducted leak segment in s1-s6 without any confusion.

In 6<sup>th</sup> level statistical analysis and 7<sup>th</sup> level pruning, the leak segment in s1-s6 was concluded. Subsequently, the leak location in s1-s6 was determined by 8<sup>th</sup> level pinpointing. As shown in Figures 4.23 (a) and (b), both results of the field measurement and simulate input achieved accurate prediction with significantly low RMSEs. The APA achieved a RMSE of 3.76m and the SPA has achieved a RMSE of 0.69m in the field prediction. In the simulated case, RMSEs of 0.07m and 0.31m were achieved by APA and SPA, respectively. The localisation accuracies were appreciable in the pipe segment with a span of 53.5m. Both predictions based on the field measurement and simulated input were time-effective, which took 4.22s in the simulated case and 7.37s in the field prediction. Lastly, the localisation results based on both field measurement and simulated input were summarised in Table 4.2.

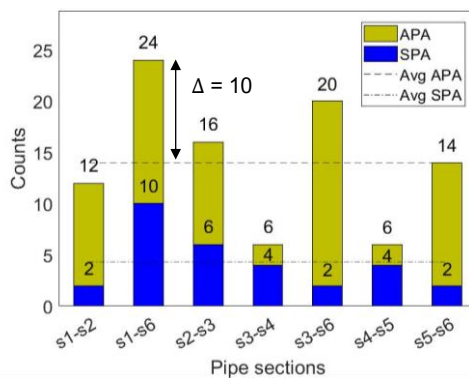


Figure 4.21 (a): 6th level general statistical analysis based on field measurement.

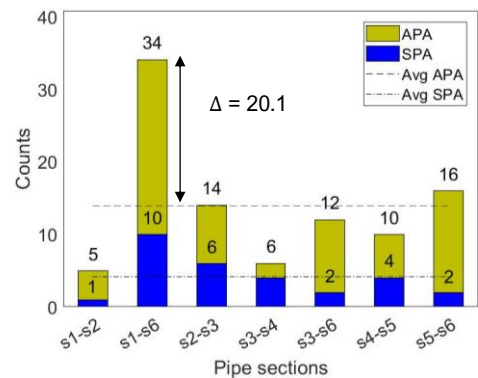


Figure 4.21 (b): 6th level general statistical analysis based on simulated input.

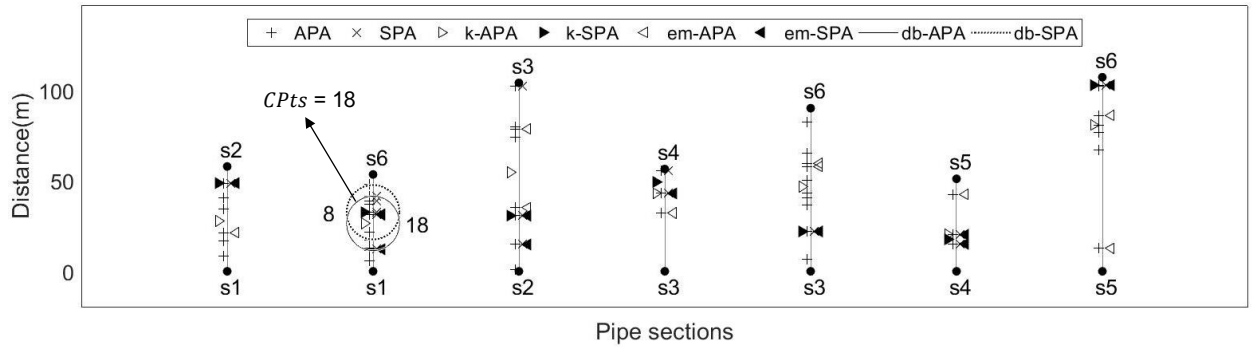


Figure 4.22 (a): 7th level pruning based on field measurement.

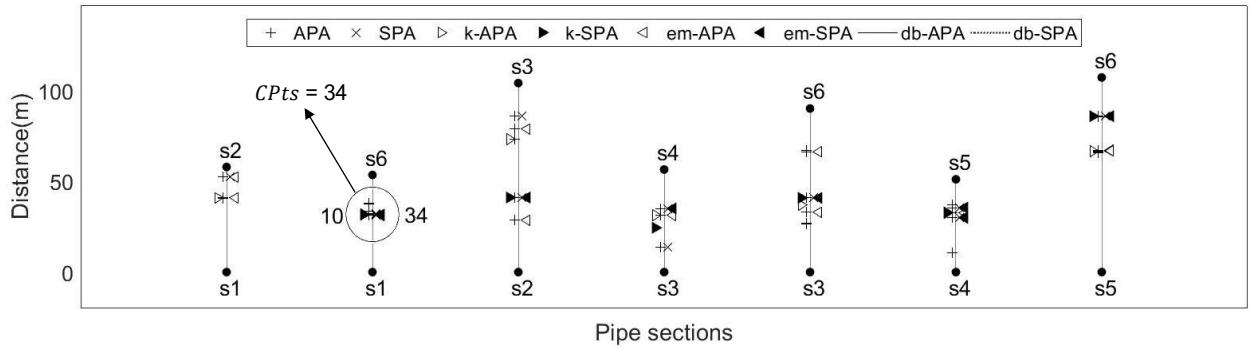


Figure 4.22 (b): 7th level pruning based on simulated input.

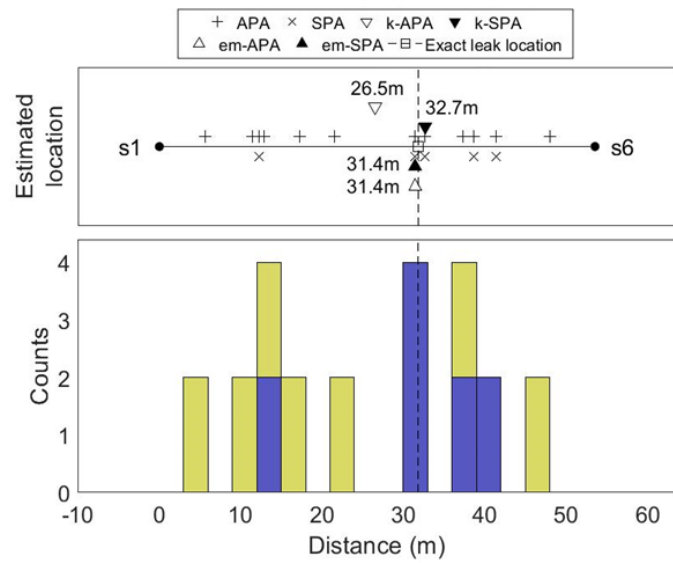


Figure 4.23 (a): 8th level pinpointing based on field measurement.



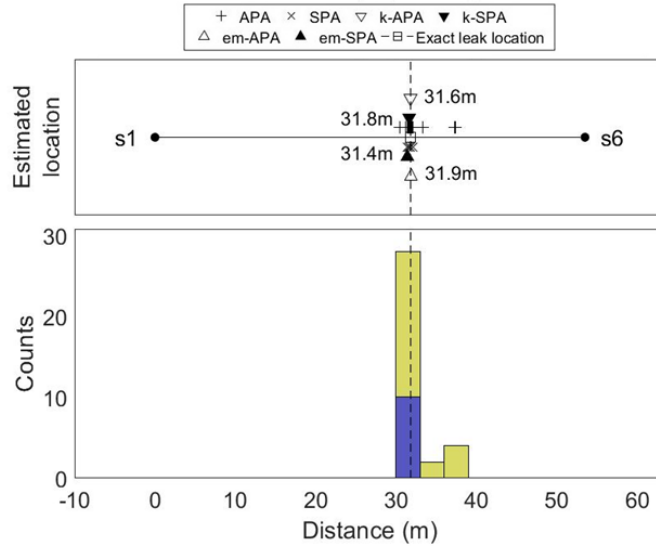


Figure 4.23 (b): 8th level pinpointing based on simulated input.

Table 4.2: Localisation results based on the field measurements and simulated inputs.

		Field measurements	Simulated inputs
$\Delta_{max}$	APA	10.0	20.1
	SPA	5.7	5.9
$Cpts_{max}$	APA	18	34
	SPA	8	10
K-mean (m)	APA	26.5	31.8
	SPA	32.7	31.6
EM-mean (m)	APA	31.4	31.9
	SPA	31.4	31.4
RMSE (m)	APA	3.76	0.07
	SPA	0.69	0.31
Time (s)		7.37	4.22

## 4.2 MDNN leak localisation

### 4.2.1 Data preparation

In the section, the MDNN was trained to predict the leak location in the same local DMA as experimented in section 4.1.4. Multiple sets of distributed leaks were simulated for the MDNN training with respect to single leak at different locations. As shown in Figure 4.24,  $n$  distributed leaks with equal separation distances,  $s$  were simulated along the 7 pipe segments of the local DMA. The TDoA of each leak location was simulated through the inverse location analysis and mathematic modelling. Table 4.3 shows the simulated dataset which consists of the input (TDoAs, receival sensors, and distance between sensors) and the targets (leak segments and locations). The training data, validation data and testing data was split with ratio 70:15:15 of the whole dataset.

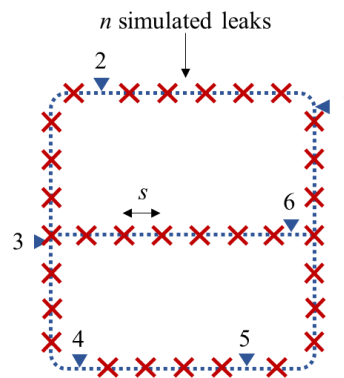


Figure 4.24: Simulated leaks in the targeted DMA.

Table 4.3: Simulated input data and respective targets.

Input (I)			Target (T)	
Receivals sensors ( <i>a-b</i> )	Distances	TDoAs	Leak segments ( <i>c-d</i> )	Leak locations
1-2	$L_{12}$	$\tau_1, \tau_2, \dots, \tau_a$	1-2	$l_1, l_2, l_3, \dots, l_a$
1-3	$L_{13}$	$\tau_1, \tau_2, \dots, \tau_a,$ $\tau_1, \tau_2, \dots, \tau_b$	1-2, 2-3	$l_1, l_2, l_3, \dots, l_a$ $l_1, l_2, l_3, \dots, l_b$
⋮	⋮	⋮	⋮	⋮
1-6	$L_{36}$	$\tau_1, \tau_2, \dots, \tau_a,$ $\tau_1, \tau_2, \dots, \tau_b,$ $\tau_1, \tau_2, \dots, \tau_c,$ $\tau_1, \tau_2, \dots, \tau_d$ $\tau_1, \tau_2, \dots, \tau_e$ $\tau_1, \tau_2, \dots, \tau_f$	1-2, 2-3, 3-4, 4-5, 5-6, 1-6	$l_1, l_2, l_3, \dots, l_a$ $l_1, l_2, l_3, \dots, l_b$ $l_1, l_2, l_3, \dots, l_c$ $l_1, l_2, l_3, \dots, l_d$ $l_1, l_2, l_3, \dots, l_e$ $l_1, l_2, l_3, \dots, l_f$
⋮	⋮	⋮	⋮	⋮
5-6	$L_{56}$	$\tau_1, \tau_2, \dots, \tau_e$	5-6	$l_1, l_2, l_3, \dots, l_e$

### Training based on different tensor shapes

Modifications of tensor shapes and dropouts during the training are the most fundamental ways to determine the optimal model of a neural network. The key ideas are to constrain the network complexity by changing the number of hidden layers and randomly drop some connections from each layer during the training. Figures 4.25, 4.46, 4.47, and 4.48 illustrated The MDNN with tensor shapes 1, 2, 3, and 4, respectively. The tensor shapes were varied by the number of hidden layers, number of dropout layers, and the size of layer. A consistent 32 samples batch size and training data of sample size,  $n = 3525$  ( $s =$

1m) were used in the training to determine the optimal tensor shape of the MDNN.

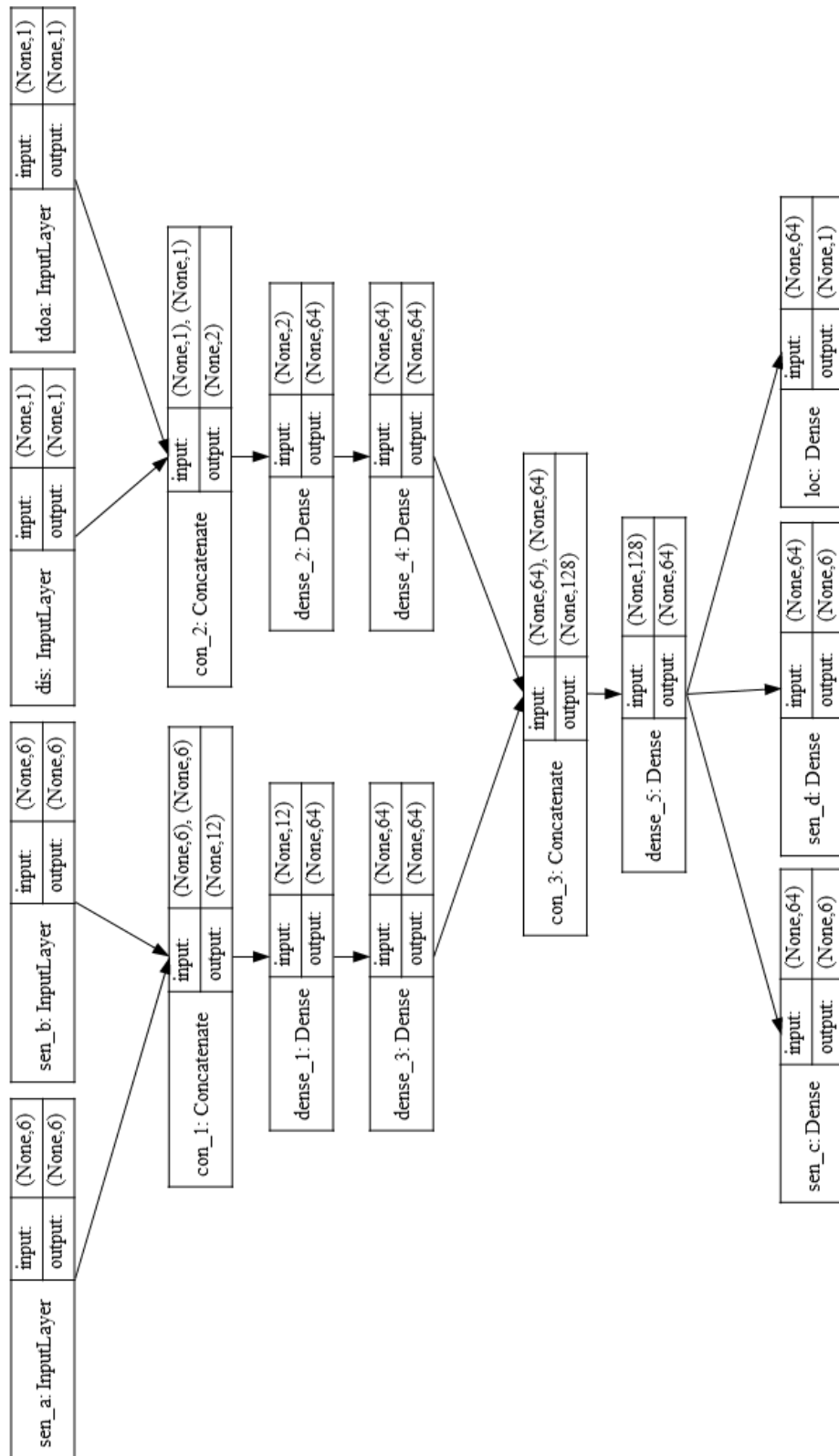


Figure 4.25: Tensor shape 1.

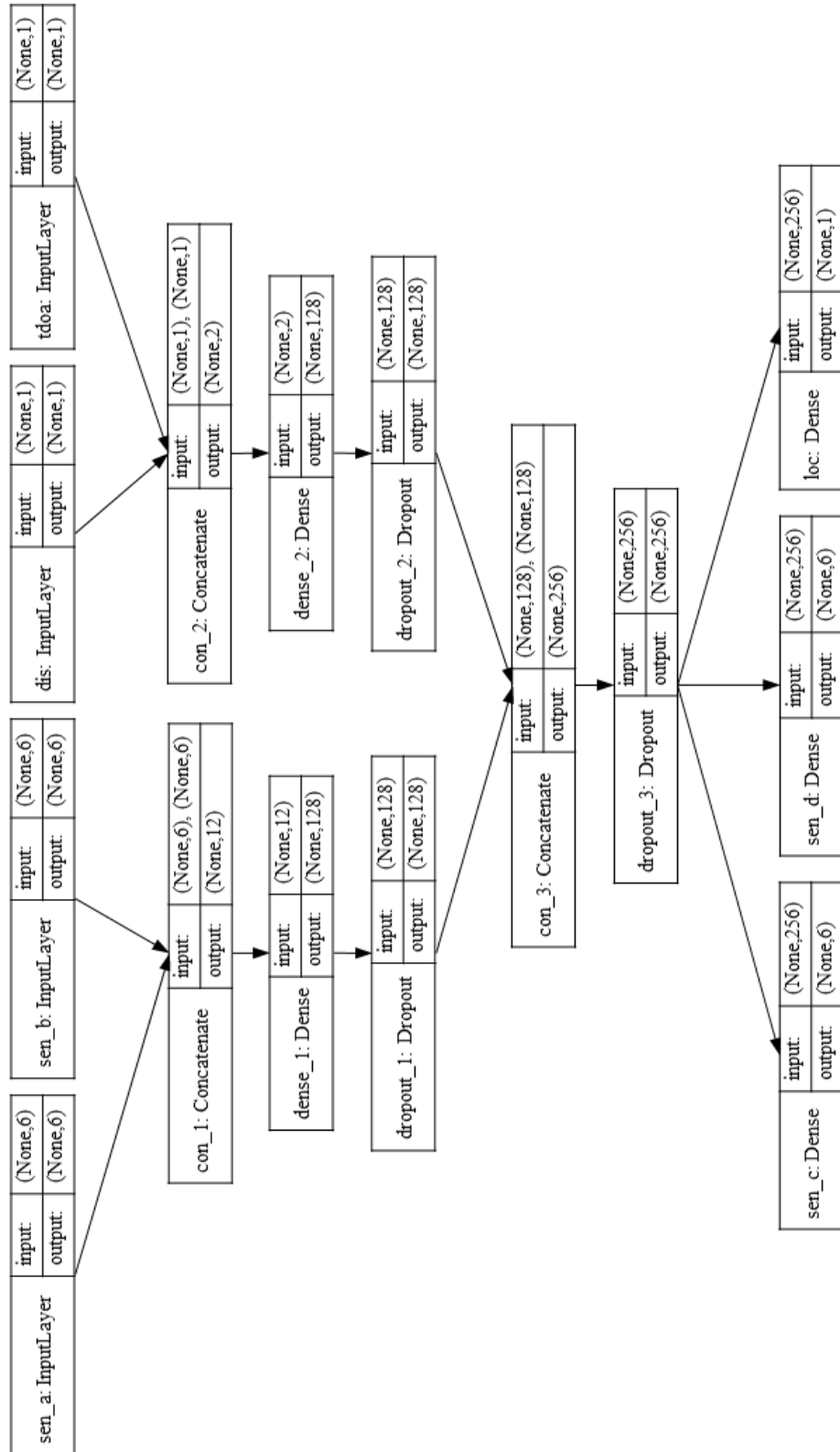


Figure 4.26: Tensor shape 2.

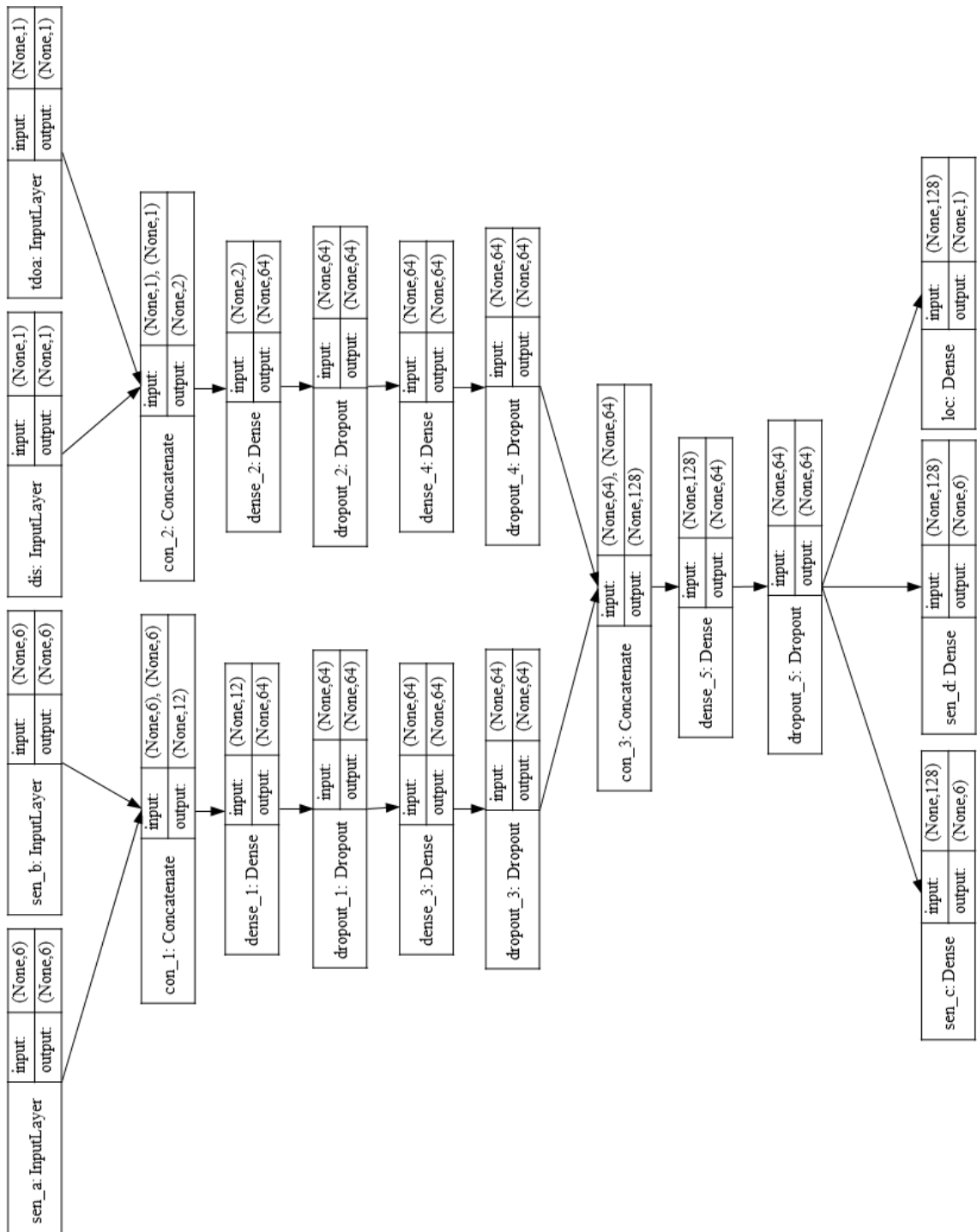


Figure 4.27: Tensor shape 3.

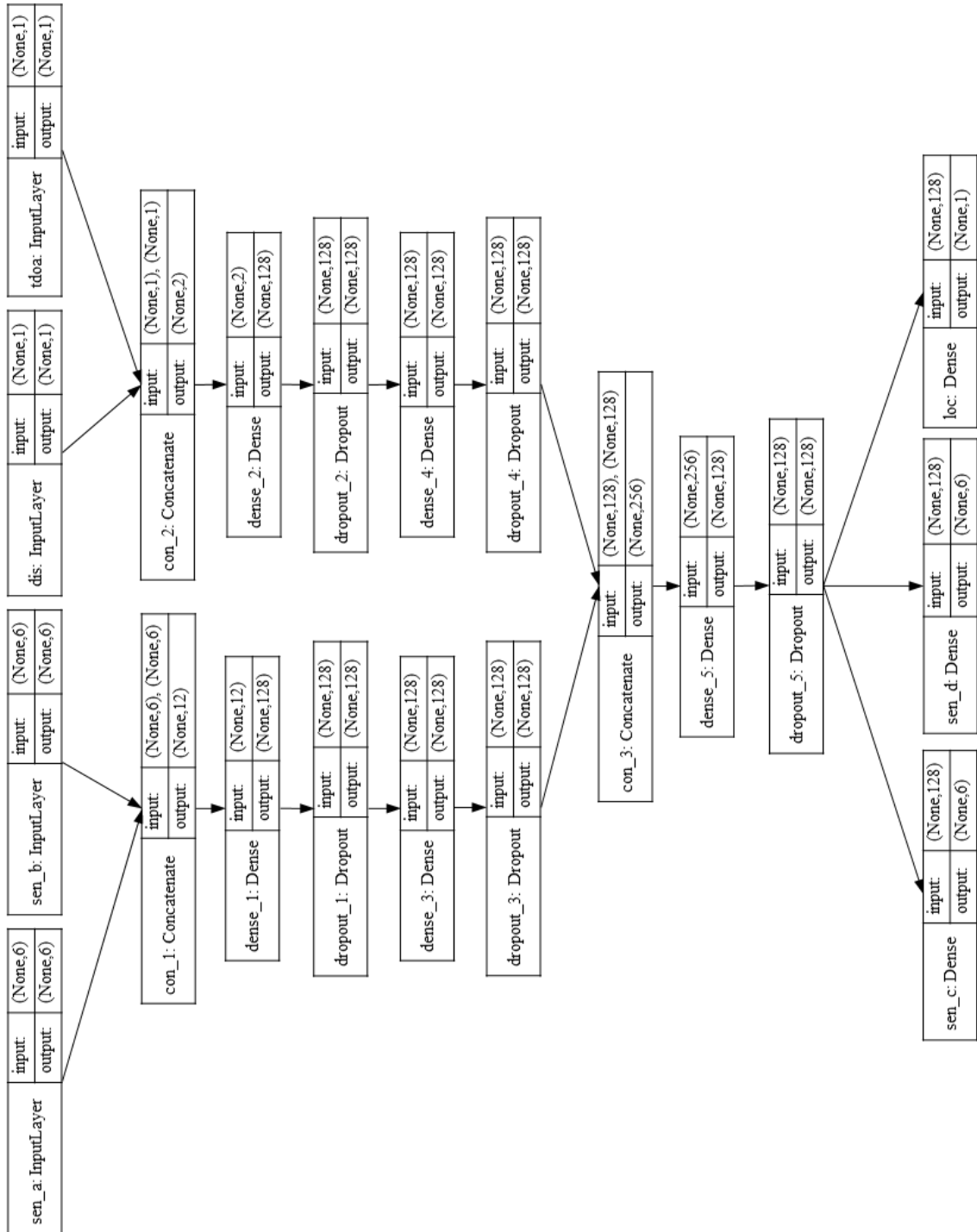


Figure 4.28: Tensor shape 4.

Figures 4.29, 4.30, 4.31, and 4.32 presented the results of training corresponding to the four tensor shapes. MAPE, accuracies, and the training losses were tabulated in Table 4.4 for ease of comparison. The training results with the higher accuracies, lower MAPE and training losses are preferable. In

general, tensor shape 2 showed the best training results and highest learning rate (least epochs needed). The tensor shape 2 fitted the training fastest, which was around 500 epochs while other tensor shapes took around 3000 epochs to achieve the equivalent accuracies. Thus, the tensor shape 2 was concluded as the optimal tensor shape of the MDNN.

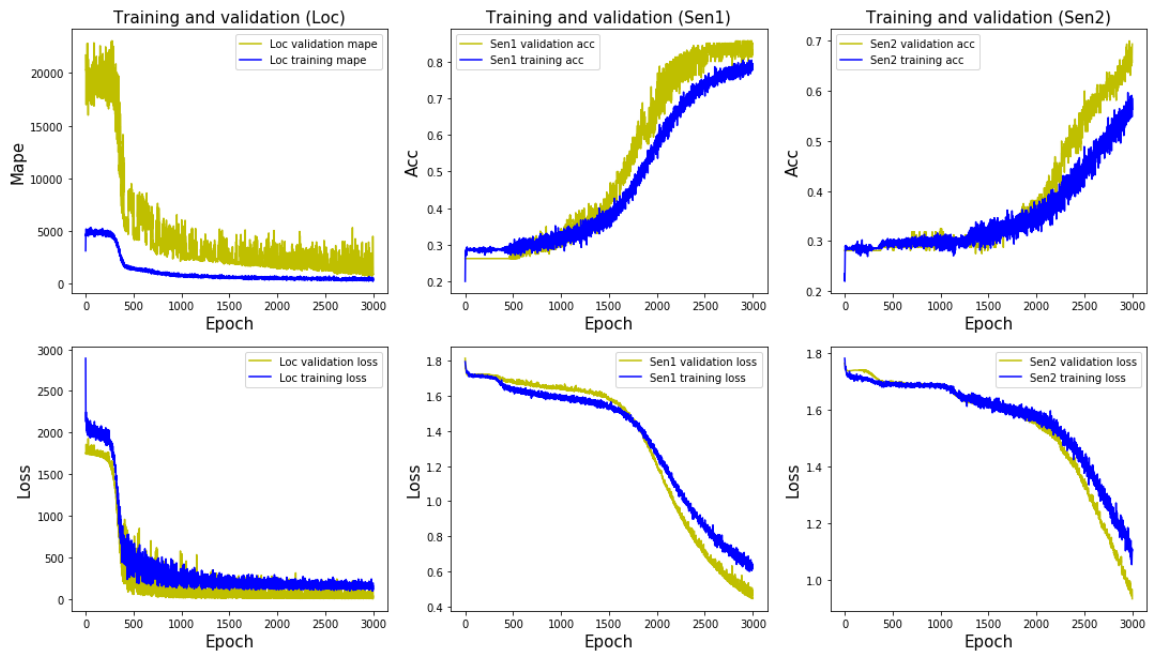


Figure 4.29: Training results of tensor shape 1.

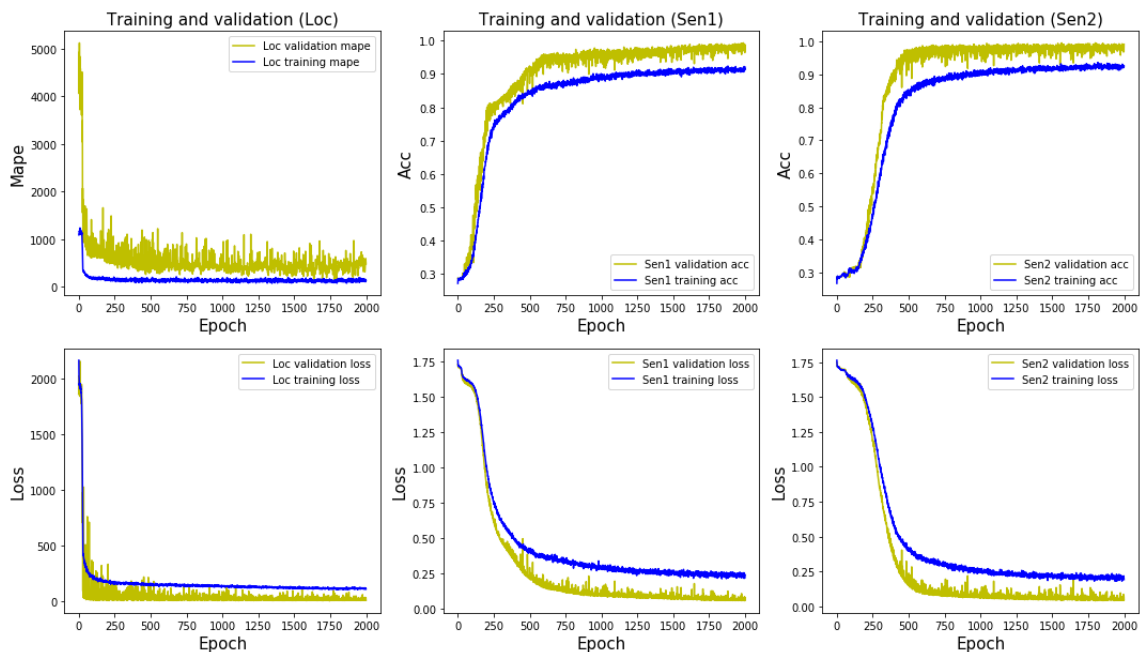


Figure 4.30: Training results of tensor shape 2.



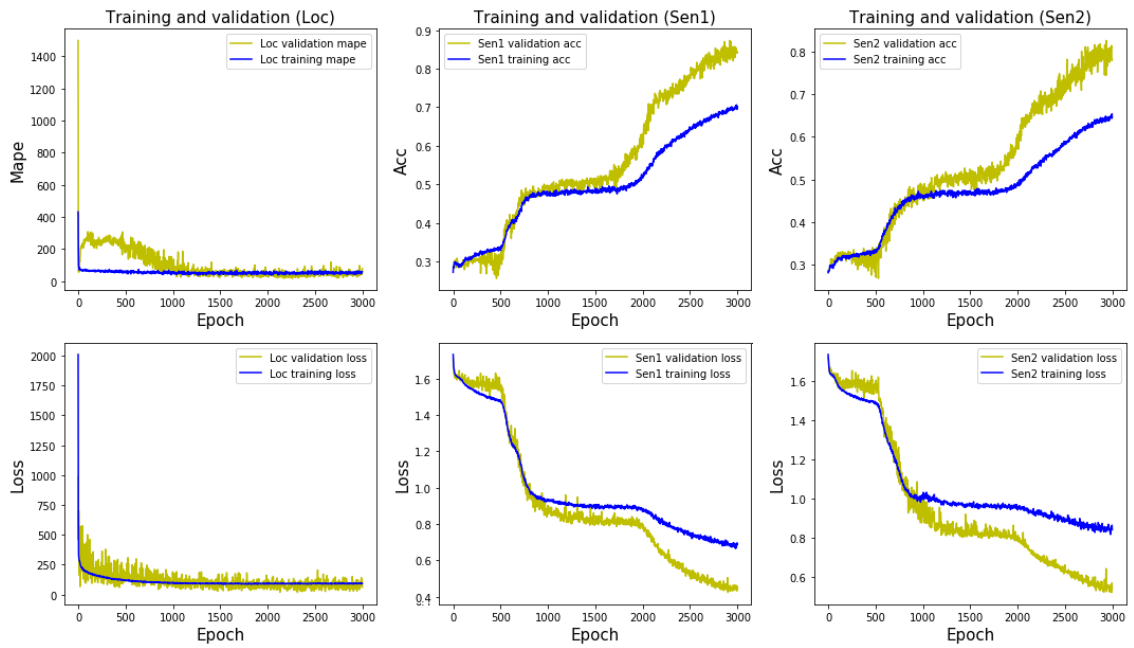


Figure 4.31: Training results of tensor shape 3.

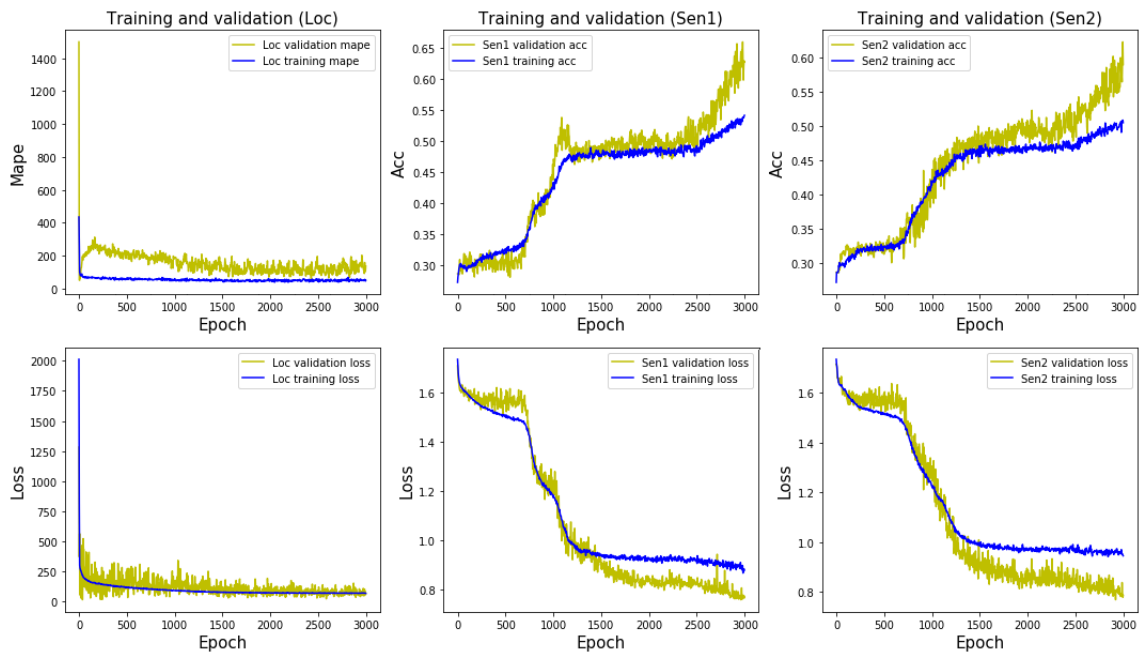


Figure 4.32: Training results of tensor shape 4.

Table 4.4: Training results based on tensor shapes 1, 2, 3, and 4.

Tensor shape	Epochs		MAPE (%)	Accuracy		Training loss		
			Leak location	Leak segment		Leak location	Leak segment	
				<i>c</i>	<i>d</i>		<i>c</i>	<i>d</i>
1	3000	T	377.2	0.78	0.56	184.5	0.62	1.10
		V	855.7	0.85	0.69	31.9	0.44	0.93
2	2000	T	139.6	0.92	0.93	108.7	0.22	0.19
		V	486.4	0.98	0.92	130.7	0.16	0.18
3	3000	T	146.0	0.69	0.65	171.9	0.78	0.95
		V	139.1	0.84	0.79	160.9	0.46	0.57
4	3000	T	192.5	0.52	0.48	146.4	0.91	0.98
		V	128.6	0.62	0.61	137.5	0.81	0.81

#### 4.2.2 Training based on different batch sizes

During the training, a dataset can always be divided into one or more batches. The batch size indicates the number of samples to work through in each iteration during a single training epoch. The popular batch sizes in neural network training are 32, 64, and 128 samples. In this section, the optimal batch size of the MDNN (tensor shape 2) was determined through training with a consistent training data of sample size,  $n = 3525$  ( $s = 1m$ ).

The training results based on 32, 64, and 128 samples were plotted in Figures 4.33, 4.34, and 4.35, respectively. The MAPE, accuracies, and the training losses were tabulated in Table 4.5. An overall improvement can be observed in the training with larger batch sizes. Both MAPE and training loss of the leak location reduced significantly as the batch size increased from 32 samples to 128 samples. Whereas the accuracies and training losses of the leak segments did not vary significantly with the batch sizes. However, the learning

rate degraded slightly as the training fitted slower when larger batch sizes were applied. 500, 750, and 1250 epochs were required by the batch sizes 32, 64, and 128 to achieve the equivalent fitting. Despite of the low learning rate, 128 samples was concluded to be the best batch size for the MDNN training due to the lowest MAPE.

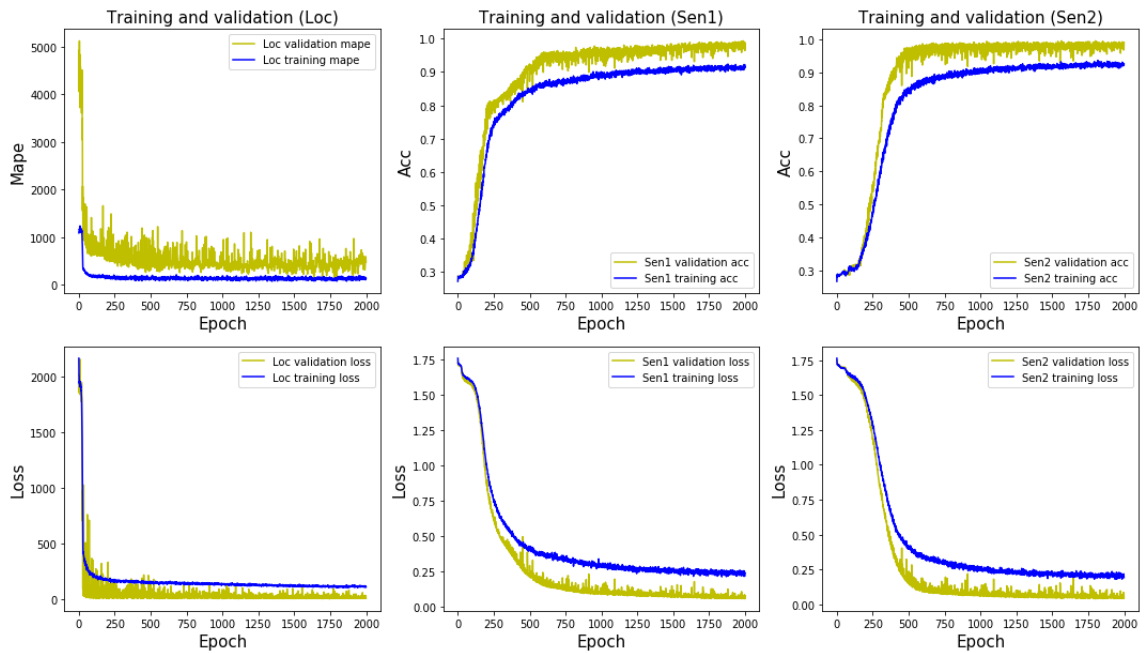


Figure 4.33: Training results based on 32 samples batch size.

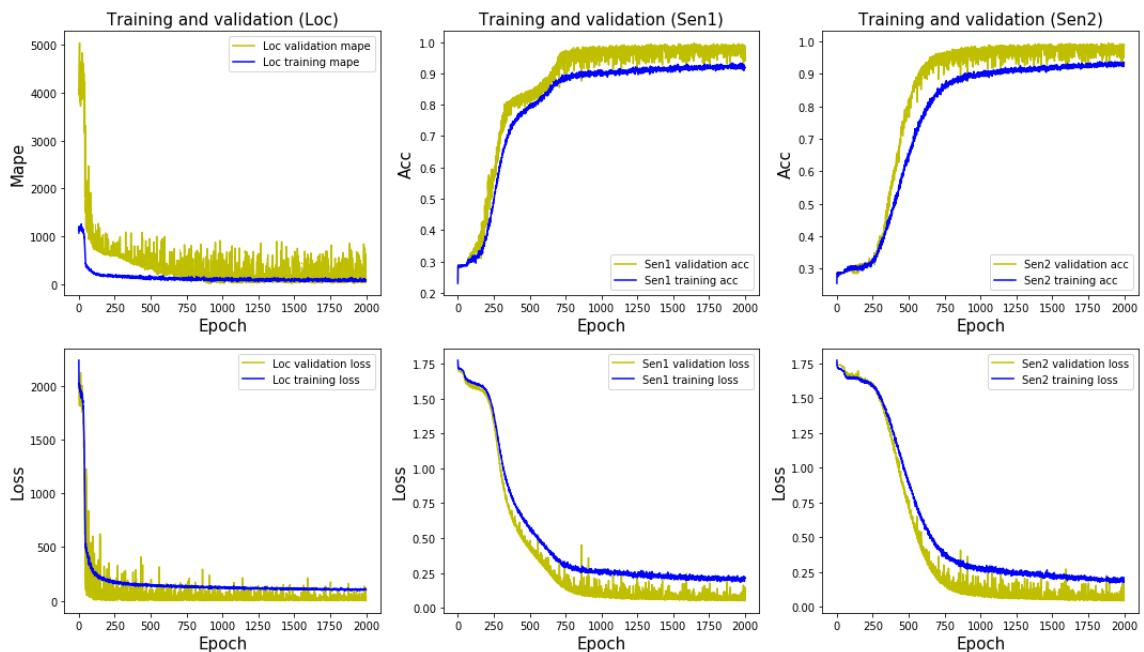


Figure 4.34: Training results based on 64 samples batch size.

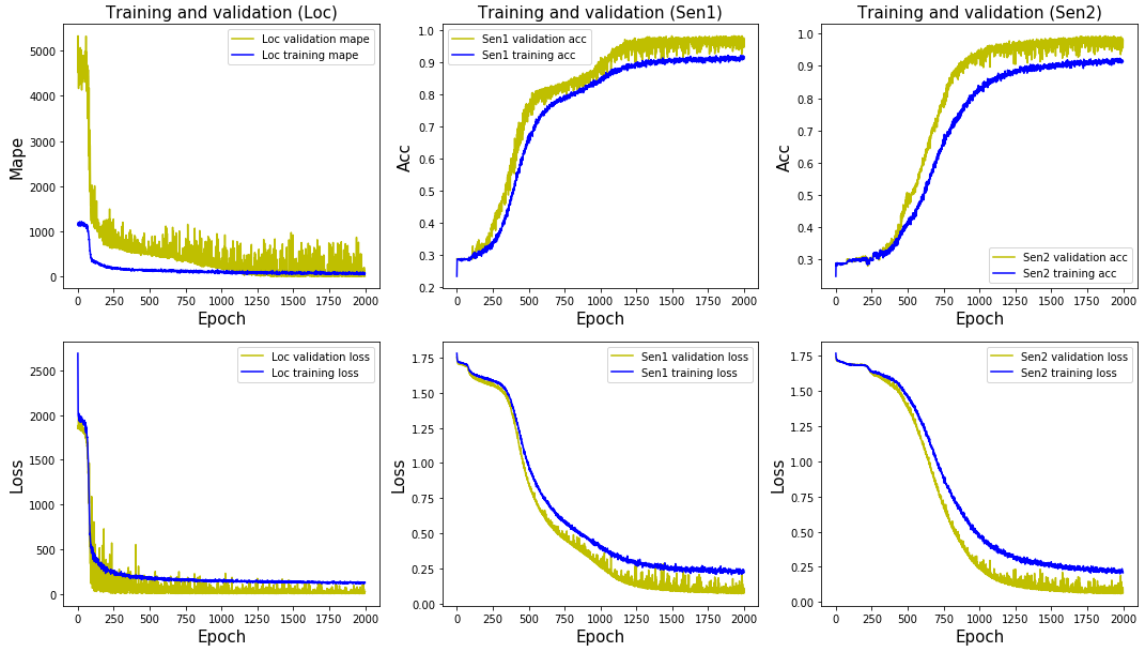


Figure 4.35: Training results based on 128 samples batch size.

Table 4.5: Training results based on batch sizes 32, 64, and 128.

Batch sizes	Epochs		MAPE (%)		Accuracy		Training loss		
			Leak location	Leak segment	<i>c</i>	<i>d</i>	Leak location	<i>c</i>	<i>d</i>
32	2000	T	139.6	0.92	0.93	108.7	0.22	0.19	
		V	486.4	0.98	0.92	130.7	0.16	0.18	
64	2000	T	77.4	0.92	0.94	103.0	0.22	0.18	
		V	49.7	0.94	0.95	99.8	0.15	0.16	
128	2000	T	72.6	0.92	0.92	94.4	0.22	0.21	
		V	14.3	0.97	0.98	23.7	0.08	0.07	

### 4.2.3 Training based on different sample sizes

In this section, the sample size of the training data was varied by  $n = 3525, 7050, 35254, 70508, \text{ and } 352540$  with  $s = 1\text{m}, 0.5\text{m}, 0.1\text{m}, 0.05\text{m}, \text{ and } 0.01\text{m}$ , respectively. The results of the corresponding training were presented in

Figures 4.36, 4.37, 4.38, 4.39, and 4.40. In overall, the training with larger sample size achieved earlier fitting. Besides, the number of epochs to complete the training had significantly reduced from 2000 to the minimum 250 as the sample size varied from 3525 to the maximum 352540. Training results based on the five sample sizes were summarised in Table 4.6. Noticeable decrements of the MAPEs and training losses with increasing sample size can be observed in Table 4.6. The best training result was achieved by the training data of sample size 352540, with MAPE of 8.7% and leak segment accuracies of 0.99.

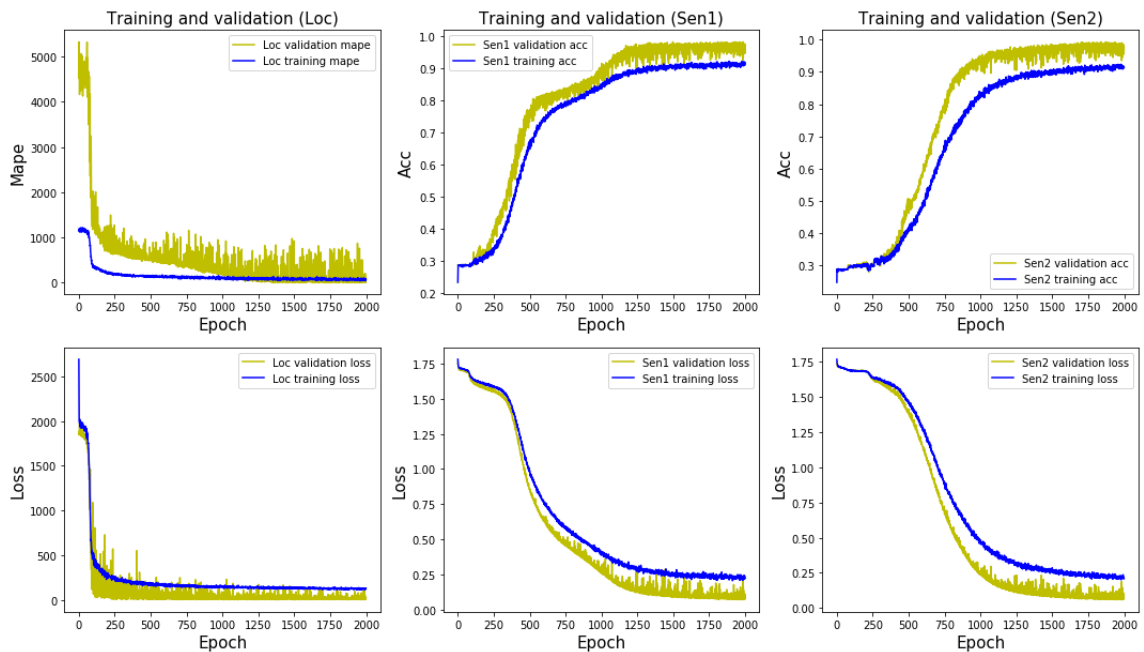


Figure 4.36: Training results based on 3525 sample size.

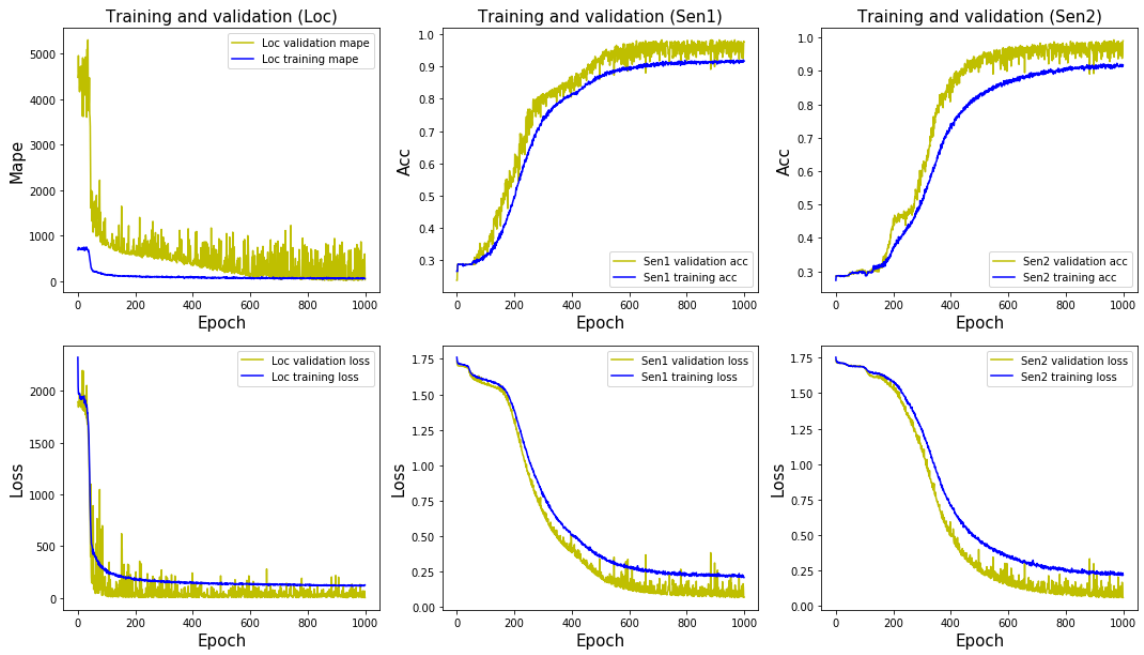


Figure 4.37: Training results based on 7050 sample size.

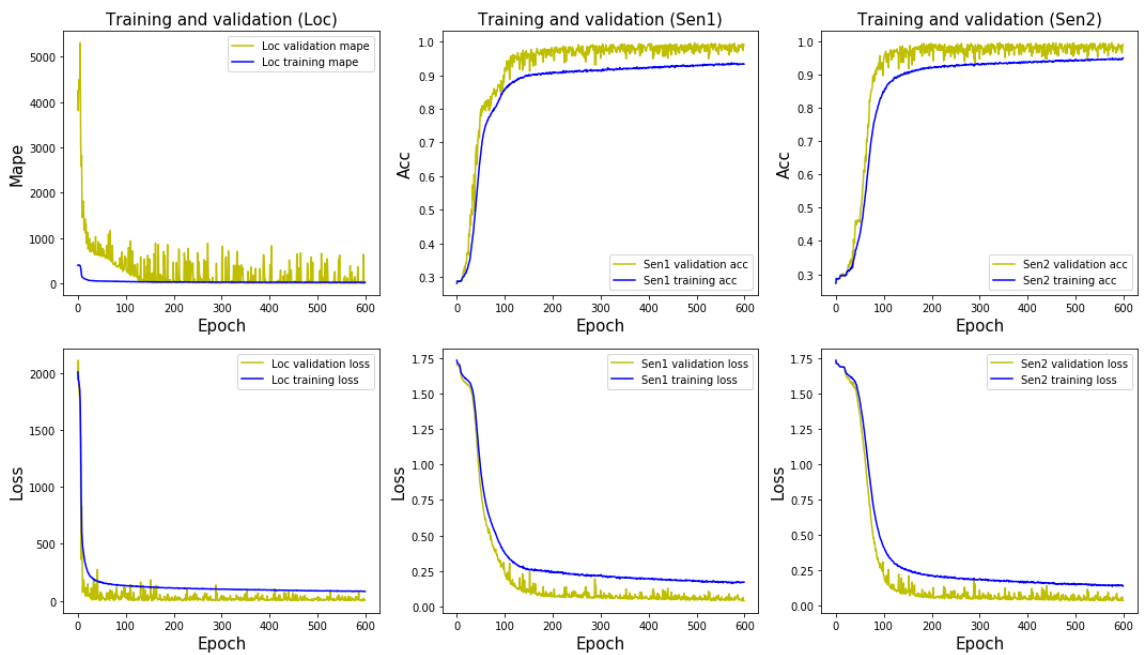


Figure 4.38: Training results based on 35254 sample size.

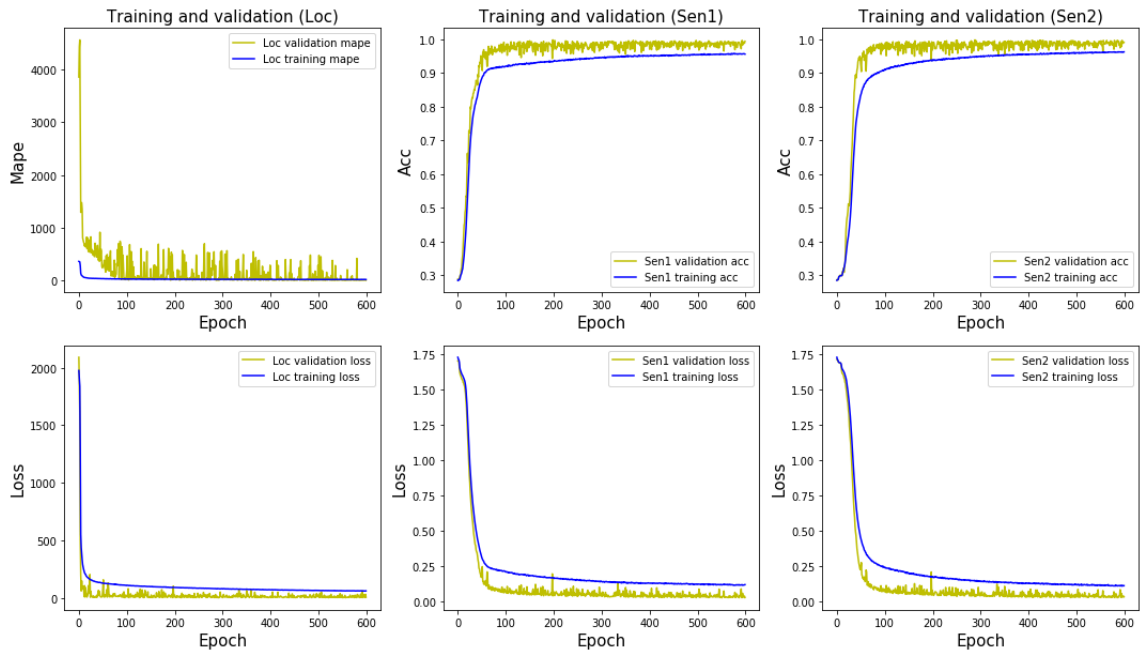


Figure 4.39: Training results based on 70508 sample size.

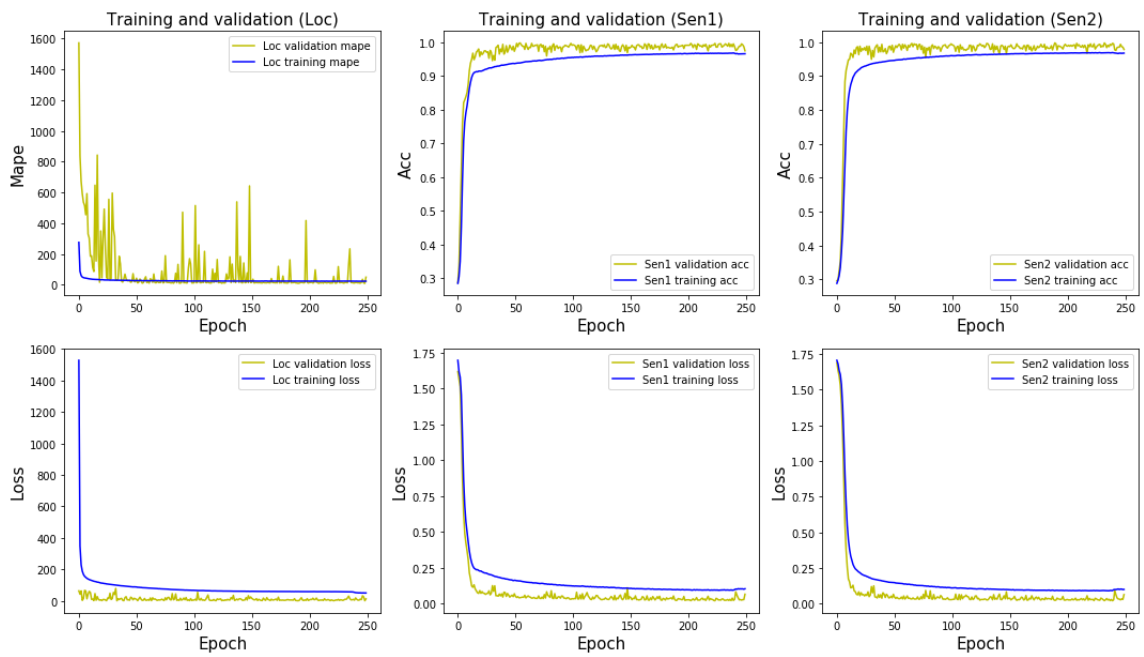


Figure 4.40: Training results based on 352540 sample size.

Table 4.6: Training results based on the five sample sizes.

Sample size	Epochs		MAPE (%)	Accuracy		Training loss		
			Leak location	Leak segment		Leak location	Leak segment	
				<i>c</i>	<i>d</i>		<i>c</i>	<i>d</i>
3525	2000	T	72.6	0.92	0.92	94.4	0.22	0.21
		V	14.3	0.97	0.98	23.7	0.08	0.07
7050	1000	T	51.8	0.92	0.92	124.4	0.21	0.22
		V	26.4	0.98	0.99	5.99	0.07	0.06
35254	600	T	29.9	0.93	0.95	82.9	0.17	0.14
		V	11.5	0.99	0.99	5.6	0.04	0.03
70503	600	T	23.3	0.96	0.96	60.4	0.12	0.11
		V	10.6	0.99	0.99	6.19	0.03	0.03
352540	250	T	19.9	0.97	0.97	45.2	0.11	0.11
		V	8.7	0.99	0.99	2.3	0.04	0.06

In the previous training, the optimal model of MDNN was identified with optimal tensor shape 2, 128 samples batch size and the training data of sample size 352540. A performance test was conducted based on the testing data and the result was presented in Table 4. In the performance test, the optimal model achieved accurate predictions in both the leak segment and location, with an average percentage error of 0.2.



Table 4.7: Results of the performance test.

Input distance (m)	Input TDoA (s)	Receival sensor		Actual leak location (m)	Predicted leak location (m)	Predicted leak segment		Percentage error (%)
		<i>a</i>	<i>b</i>			<i>c</i>	<i>d</i>	
143.6	-0.0668	1	3	25.05	25.1	1	6	0.2
200.1	-0.0357	1	4	75.05	75.3	6	3	0.3
104.1	0.0686	2	3	100.05	100.1	2	3	0.0
211.7	0.0989	2	5	175.05	175.2	4	5	0.1
143.6	0.0046	3	1	75.05	75.1	3	6	0.1
107.6	-0.0411	3	5	25.05	25.1	3	4	0.2
107.6	0.0304	3	5	75.05	75.3	4	5	0.3
200.1	-0.0357	4	1	75.05	75.5	3	6	0.6
160.6	-0.0432	4	2	50.05	50.3	4	3	0.5
146.6	0.0739	4	6	125.05	125.2	3	6	0.1
160.7	0.0281	5	1	100.05	100.0	5	6	0.0
211.7	0.0632	5	2	150.05	150.2	3	2	0.1
111.4	0.0276	6	2	75.05	75.4	1	2	0.5
146.6	0.0025	6	4	75.05	75.1	6	3	0.1
								0.2

#### 4.2.4 Prediction based on the field measurements

In this section, the optimal model of MDNN was used to predict the actual leak location in the local DMA. Field measurements were fed into the MDNN model as the input features. Table 4.8 presented the results of field prediction where the predicted leak locations and the actual leak locations were compared in term of percentage error. The overall result shows that the MDNN had accurately predicted the leak segment 1-6 and achieved an average 3.2% leak location error. The deviations between the predictions and the actual leak

locations were plotted in Figure 4.41 for results visualisation. The results of field prediction were excellent based on the optimal model of MDNN.

Table 4.8: Results of the field prediction.

Input distance (m)	Input TDoAs (s)	Receival sensor		Actual leak location (m)	Predicted leak location (m)	Predicted leak segment		Percentage error (%)
		<i>a</i>	<i>b</i>			<i>c</i>	<i>d</i>	
143.6	-0.0851	1	3	31.8	32.4	1	6	1.9
200.1	-0.0981	1	4	31.8	30.5	1	6	4.1
160.7	-0.0596	1	5	31.8	32.5	1	6	2.2
53.5	0.0084	1	6	31.8	32.2	1	6	1.2
111.4	0.0623	2	6	89.7	91.8	1	6	2.3
143.6	0.0851	3	1	102.8	107.8	6	1	4.9
200.1	0.0981	4	1	168.3	165.2	6	1	1.8
160.7	0.0596	5	1	128.9	119.3	6	1	7.4
								3.2

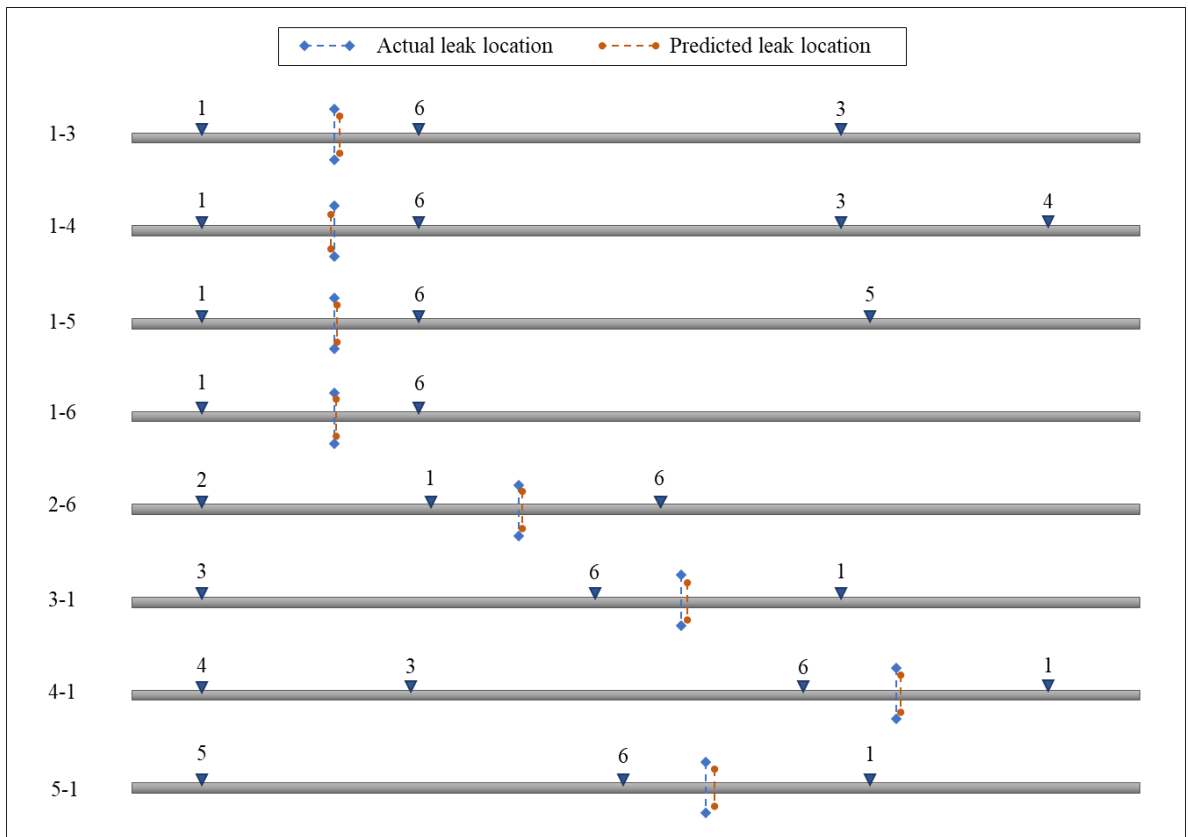


Figure 4.41: Results of the field prediction.

### 4.3 Comparisons of the MLAF and MDNN

In this section, the performances of the MLAF and MDNN were compared in terms of the localisation accuracies, time taken for the localisation process, and flexibility of application.

#### 4.3.1 Localisation accuracies

The localisation accuracies of both MLAF and MDNN have been validated in the field prediction. The MLAF has achieved the best RMSE of 0.69m in the APA localisation, which was 2.2% location error. The location error was slightly better than the MDNN which was 3.2%. In overall, both the localisation accuracies were excellent in the pipe segment of 31.8m span.

### **4.3.2 Localisation times**

Both localisation processes of the MLAF and MDNN were time-effectively. The localisation time of the MLAF during the field test was 7.37s, whereas the localisation time of the MDNN was less than 1s. However, the MDNN required a pre-training before the localisation and took time to determine the optimal model. The training durations were varied with the sample size of the training data, the time ranged from 15 minutes to more than 3 hours in the previous trials. The optimal model of MDNN took 3.5 hours to complete the training, which is less practical for the application of adaptive localisation system.

### **4.3.3 Flexibilities of application**

The high adaptiveness of MLAF in various topological different piping networks has been proven in the previous characterisation tests. Thus, the flexible in application and can be widely implemented in various topological different piping networks. On the other hand, the MDNN is suitable for a fixed monitoring system in a targeted piping network. However, the MDNN can always be trained for localisation in different piping networks. In other words, the MDNN can be applied to any piping network, but requires pre-application training.

## CHAPTER 5

### CONCLUSION AND RECOMMENDATIONS

#### 5.1 Conclusion

In the past decades, the conventional acoustic methods have been well-engineered to localise the leak in the single pipeline system. In this thesis, two intelligent leak localisation systems based on monogenous sensing principle of acoustic devices were developed to localise a single leak in the WDNs.

In the first methodology, the MLAF was formulated for adaptive leak localisation. The multi-directional waves issue and the analytical complexity in piping networks were solved by the multi-level hierarchical analytics and reasoning process. The overall MLAF localisation processes were prompt and seamless by adopting an automated flow control system. The accuracies and adaptiveness of the MLAF has been validated through multiple sets of characterisation tests in respect of different network topologies. A field prediction in a local DMA with excellent localisation result of 2.2% location error further affirmed the feasibility of the MLAF in the real application of localising leak.

In the second methodology, the MDNN was developed based on the Keras module. The mixed-model neural networks architecture of the MDNN has achieved simultaneous predictions of both leak segment and location. The prediction was based on aggregated regression and classification sequential learnings of input features such as the TDoAs, receiver sensors, and the distances between sensors. The MDNN was trained with different tensor shapes, batch sizes and sample sizes of training data to determine the optimal model. The optimal model of MDNN was validated with 2.3% MAPE and 0.99 training accuracies. Finally, the prediction of the optimal model was validated with 3.2% location error in the field prediction.

Finally, both MLAF and MDNN have achieved time-effective and accurate localisations in the WDNs. The adaptiveness of the MLAF has been well-proven through a series of characterisation tests in respect of the topological variation. The MDNN on the other hand, can be flexibly applied for different piping networks with a pre-application training. However, the MDNN is slightly inferior to the MLAF in term of the adaptiveness. In conclusion, the MLAF fitted perfectly with the high portability of acoustic sensing system to cater for flexible leak localisation. Whereas the MDNN is suitable for a long-term monitoring where the sensing system is fixed in one specific DMA.

## **5.2 Recommendations for future works**

The capabilities to localise single leak in piping networks of the proposed leak localisation systems have been verified with high accuracies. In the future works, it is of interest to further develop the leak localisation systems

for multiple leaks localisation. Some possible future research directions in respect of the multiple leaks localisation are presented as follows:

- The location analyses and reasoning processes of the MLAF can be further modified to cater the multiple leaks localisation. Thresholds can be characterised and set in the cluster analyses to prune the multiple leak segments.
- A method to test and validate the multiple leaks localisation shall be formulated as an alternative for the field experiments.
- The mathematical modelling can be further improved based on multiple sources superposition theorem. Besides, a deliberate way of simulating possible noises and paths damping will be better to mimic the real leak cases for both characterisation and training purposes.
- Modification of mixed-model architecture with more layers for leak locations differentiation can be considered. However, the modification depends on the capability of the present model in predicting multiple leaks.

Furthermore, proceeding works of the present methods can be considered. If the MDNN is possibly implemented for long term monitoring in a fixed DMA, the training can be fed with the real field measurements instead of the simulated data for a better prediction. Besides, more field predictions can be conducted by further accessing more real DMA with single and multiple leaks in the future.

## REFERENCES

- Abdulshaheed, A., Mustapha, F. & Anuar, M., 2017. Pipe material effect on water network leak detection using a pressure residual vector method. *J. Water Resour. Plann. Manage*, 144(4).
- Afifi, M., Nordin, S. & Vijanth, A., 2011. *Pressure point analysis for early detection system*. Penang, Malaysia, March.
- Afzal, M. & Udpa, S., 2002. Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. *NDT & E International*, 35(7), p. 449–457.
- Agathokleous, A., Xanthos, S. & Christodoulou, S., 2015. Real-time monitoring of water distribution networks. *Water Utility Journal*, Volume 10, pp. 15-24.
- Allen, M., Preis, A., Iqbal, M. & Whittle, A. J., 2013. *Water distribution system monitoring and decision support using a wireless sensor network*. Hawaii, United States, July.
- Anon., 1999. *Technical review of leak detection technologies, vol. i: Crude oil transmission pipelines*, Alaska: Alaska Department of Environmental Conservation.
- Anon., 2002. *Trenchless technology network, underground mapping, pipeline location technology and condition assessment*, Birmingham: Infrastructure Engineering and Management Research Centre, The University of Birmingham.



- Aya, A., Oussama, G., Obeid, A. & Mohamed, A., 2017. *Leak detection in water pipeline by means of pressure measurements for WSN*. Fez, Morocco, May.
- Bakker, M., Van Schagen, K. M. & Timmer, J., 2003. Flow control by prediction of water demand. *Journal of Water Supply. Research and Technology-AQUA*, 52(6), pp. 417-424.
- Butterfield, J. D. et al., 2017. Prediction of leak flow rate in plastic water distribution pipes using vibro-acoustic measurements. *Structural Health Monitoring*, 17(4), pp. 959-970.
- Caputo, A. & Palegagge, P., 2003. Using neural networks to monitor piping systems. *Process Safety Progress*, 22(2), pp. 119-127.
- Choi, D., Kim, S., Choi, M. & Geem, Z., 2016. Adaptive kalman filter based on adjustable sampling interval in burst detection for water distribution system. *Water*, 8(4), p. 142.
- Clair, A. & Sinha, S., 2014. Development of a standard data structure for predicting the remaining physical life and consequence of failure of water pipes. *J. Perform. Constr. Facil*, Volume 28, pp. 191-203.
- Cugueró-Escofet, M. A., Puig, V. & Quevedo, J., 2017. Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index. *Application to the Barcelona water network. Control Engineering Practice*, Issue 63, pp. 1-12.
- Dong, Y. & Yu, D., 2005. Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis. *J. Loss Prev. Process Ind*, Volume 18, pp. 83-88.

Eiswirth, M., Heske, C., Burn, L. S. & DeSilva, D., 2001. *New methods for water pipeline assessment*. Berlin, Germany, Proceedings of the 2nd World Water Congress of the International Water Association.

Ferziger, J. H. & Peric, M., 2002. *Computational Methods for Fluid Dynamics*. 3rd ed. New York: Springer.

Foo, L. et al., 2018. *Feasibility of cross-time-frequency spectrum based leakage location in gas pipeline using fiber taper sensor*. Jeju, TENCON 2018.

Foo, L. et al., 2018. *Feasibility of cross-time-frequency spectrum based leakage location in gas pipeline using fiber taper sensor*. Jeju, Korea, TENCON 2018.

Gao, Y. et al., 2004. A model of the correlation function of leak noise in buried plastic pipes. *Journal of Sound and Vibration*, 277(1-2), p. 33–148.

Ghazali, M. et al., 2010. Instantaneous phase and frequency for the detection of leaks and features in a pipeline system. *Structural Health Monitoring*, 10(4), pp. 351-360.

Guatam, A., Kumar, A., Ritu, R. S. & Priye, V., 2016. Optical sensing and monitoring architecture for pipelines using optical heterodyning and FBG filter. *International Journal for Light and Electron Optics*, 127(20).

Hidir, A. R., 2018. *Xavier: Reducing non-revenue water by 1pc will cost RM800,000*. [Online]

Available at: <https://www.nst.com.my/news/nation/2018/11/435434/xavier-reducing-non-revenue-water-1pc-will-cost-rm800000>

[Accessed 27 2019].

Hunaidi, .. & Wang, A., 2004. *Acoustic methods for locating leaks in municipal water pipe networks*. Jordan, International Water Demand Management Conference.

Hunaidi, O. & Wang, A., 2004. *Acoustic methods for locating leaks in municipal water pipe networks*. Jordan, International Water Demand Management Conference.

Hunaidi, O., Wang, A., Bracken, M. & Gambino, T., 2004. *Acoustic methods for locating leaks in municipal water pipe networks*. Dead Sea, Jordan, s.n.

Ide, H. & Kurita, T., 2017 . *Improvement of learning for CNN with ReLU activation by sparse regularization*. Anchorage, AK, s.n.

Jacobsz, S. W. & Jahnke, S. I., 2019. Leak detection on water pipelines in unsaturated ground by discrete fibre optic sensing. *Structural Health Monitoring*, pp. 1-18.

Jafar, R., Shahrour, I. & Juran, I., 2010. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling*, Volume 51, pp. 1170-1180.

Jiang, T. et al., 2017. Application of FBG based sensor in pipeline safety monitoring. *Applied Sciences*, 7(6), pp. 540-545.

Jung, D. & Lansey, K., 2014. Burst detection in water distribution system using the extended kalman filter. *Procedia Engineering*, 70(12), p. 902–906.

Kayaalp, F., Zengin, A., Kara, R. & Zavrak, S., 2017. Leakage detection and localization on water transportation pipelines: a multi-label classification approach. *Neural Computing & Applications*, 28(10), pp. 2905-2914.

Khulief, Y. A. & Khalifa, E. A., 2012. *On the In-Pipe Measurements of Acoustic Signature of Leaks in Water Pipelines*. s.l., Reseach Gate.

Kim, Y. et al., 2016. Robust leak detection and its localization using interval estimation for water distribution network.. *Computers & Chemical Engineering*, 92(C), pp. 1-17.

Kleij, F. C. & Stephenson, M. J., 2002. *Acoustic logging—the bristol water experience*. Lemesos, Cyprus, Proceedings of The IWA Specialised Conference: Leakage Management—A Practical Approach, International Water Association.

Law, Z. J., Pua, C. H., Lin, H. S. & A., R. F., 2018. *Application of fiber laser dynamics in leak detection for operating water pipeline*. Limerick, Ireland, 12th International Conference on Sensing Technology (ICST).

Levenson, M. & Daley, B., 2010. *A catastrophic rupture hits regions water system*. [Online]

Available at: <http://www.boston.com/news/local/massachusetts/articles/2010/05/02/>

[Accessed 2 7 2019].

Lighthill, M., 1952. *On sound generated aerodynamically: 1. General theory*. 564-587, s.n.

Lin, W. et al., 2015. A dual-sensor-based method to recognize pipeline leakage and interference signals. *Journal of Loss Prevention in the Process Industries*, Volume 38, pp. 79-86.

Li, S. et al., 2014. Leak location in gas pipelines using cross time–frequency spectrum of leakage-induced acoustic vibrations. *Journal of Sound and Vibration*, 333(17), pp. 3889-3903.

Li, S. et al., 2014. Leak location in gas pipelines using cross-time–frequency spectrum of leakage-induced acoustic vibrations. *Journal of Sound and Vibration*,, 333(17), pp. 3889-3903.

Liu, L. & Özsu, M., 2009. Density-based Clustering. In: *Encyclopedia of Database Systems*. Boston, MA: Springer.

Li, Y., Fan, Y. & Ni, N., 2018. *A pipeline leakage locating method based on the gradient descent algorithm*. Maharashtra, India, AIP Conference Proceedings .

Lu, H., Iseley, T., Behbahani, S. & Fu, L. D., 2020. Leakage detection techniques for oil and gas pipelines: State-of-the-art. *Tunnelling and Underground Space Technology*, Volume 98.

Lusa, T. et al., 2017. Acoustic emission testing of cryogenic pipelines in operating conditions. *Advances in Acoustic Emission Technology*, Volume 179, pp. 347-355.

Makar, J. & Mcdonald, S., 1996. *Assessment of the hydroscope 201TM condition index evaluation of gray cast iron pipe from gatineau*, Ottawa, Canada: Tech. Rep. QUEBEC NRC REPORT A-7015.3, National Research Council.

Makeen, A., Ghazali, M., Malik, A. & Adnan, N., 2014. Leak detection in pipeline using wavelet and ceptrum analysis. *Australian Journal of Basic and Applied Sciences*, 8(15), pp. 353-355.

- Maninder, P., Dixon, N. & Flint, j., 2010. *Detecting and locating leaks in water distribution polyethylene pipes*. London, Proceedings of the World Congress on Engineering 2010 Vol II .
- Meseguer, J., Mirats-Tur, J., Cembrano, G. & Puig, V., 2015. Model-based monitoring techniques for leakage localization in distribution water networks. *Procedia Engineering*, Volume 119, pp. 1399-1408..
- Mohamed, T. & Hichem, F., 2018. *Fuzzy and possibilistic clustering for multiple instance linear regression*. Rio de Janeiro, 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).
- Mounce, S. & Machell, J., 2007. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal*, 3(1), p. 21–31.
- Muhlbauer, W. K., 2004. *Pipeline risk management manual: ideas, techniques, and resources*. s.l.:Gulf Professional Publishing.
- Mukhopadhyay, S. & Srivastava, G. P., 2000. Characterization of metal loss defects from magnetic flux leakage signals with discrete wavelet transform. *NDT & E International*, 33(1), p. 57–65.
- Muranho, J. et al., 2014. Pressure-dependent Demand and Leakage Modelling with an EPANET Extension. *WaterNetGen. Procedia Engineering*, Volume 89, p. 632–639.
- MWA, 2017. *Malaysia Water Industry Guide 2016*, Kuala Lumpur: Malaysian Water Association.

- Nasirian, A., Maghrebi, M. & Yazdani, S., 2013. Leakage detection in water distribution network based on a new heuristic genetic algorithm model. *Journal of Water Resource and Protection*, 5(3), p. 294–303.
- Nestleroth, J. B. & Bubenik, T. A., 1999. *Magnetic flux leakage (MFL) technology for natural gas pipeline inspection*, s.l.: Tech. Rep: The Gas Research Institute.
- Okeya, I., Kapelan, Z., Hutton, C. & Naga, D., 2014. Online burst detection in water distribution systems using the kalman filter and hydraulic modelling. *Procedia Engineering*, 89(16), p. 418–427.
- Png, W. H., 2019. *Design and development of Mach-Zehnder Interferometer fiber sensor for structural health monitoring*, Selangor: Master dissertation, Universiti Tunku Abdul Rahman.
- Png, W. H., Lin, H. S., Pua, C. H. & Rahman, F. A., 2018. Pipeline monitoring and leak detection using Loop integrated Mach Zehnder Interferometer optical fiber sensor. *Optical Fiber Technology*, Volume 46, pp. 221-225.
- Ravichandran, T. et al., 2021. Ensemble-Based Machine Learning Approach for Improved Leak Detection in Water Mains. *Journal of Hydroinformatics*, 23(2), pp. 307-323.
- Rizzo, P., 2010. Water and wastewater pipe nondestructive evaluation and health monitoring: a review. *Advances in Civil Engineering*, p. 1–13.
- Sadeghioon, A., Metje, M., Chapman, D. & Anthony, C., 2014. Smartpipes: Smart wireless sensor networks for leak detection in water pipelines. *Journal of Sensors and Actuator Networks*, 3(1), p. 6478.

Salam, A., Tola, M., Selintung, M. & Maricar, F., 2014. On-line monitoring system of water leakage detection in pipe networks with artificial intelligence. *ARPN Journal of Engineering and Applied Science*, 9(10), p. 1817–1822.

Santos, P. et al., 2011. An LPV modeling and identification approach to leakage detection in high pressure natural gas transportation networks. *Control Systems Technology, IEEE*, 19(1), pp. 77-92.

Seco, F. & Jiménez, A. R., 2012. Modelling the Generation and Propagation of Ultrasonic Signals in Cylindrical Waveguides. *Ultrasonic Waves*.

SelangorJournal, 2020. *Air Selangor: Non-revenue water reduced below 30pct in 2019, best-ever record*, Selangor: Selangor Journal.

Serveh, K., Pejman, T. & Muhammed, S., 2019. Linking Morphology of Porous Media to Their Macroscopic Permeability by Deep Learning. *Transport in Porous Media*, Volume 131, pp. 427-448.

Shinozuka, M. et al., 2010. *Nondestructive monitoring of a pipe network using a MEMS-based wireless network*. San Diego, United States, March.

Soldevila, A. et al., 2016. Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, Volume 55, pp. 162-173.

Srirangarajan, S. et al., 2013. Wavelet-based burst event detection and localization in water distribution systems. *Journal of Signal Processing Systems*, 72(1), pp. 1-16.

Sun, L. & Li, Y., 2010. *Acoustic emission sound source localization for crack in the pipeline*. Xuzhou, China, May.



- Sun, Z. et al., 2011. Magnetic induction-based wireless sensor networks for underground pipeline monitoring. *Ad Hoc Networks*, 9(3), pp. 218-227.
- Talathi, S. & Vartak, A., 2016. *Improving performance of recurrent neural network with relu nonlinearity*. San Diego, CA, s.n.
- Tan, P., Steinbach, M. & Kumar, V., 2005. Cluster Analysis: Basic Concepts and Algorithms. In: *Introduction to Data Mining*. s.l.:s.n., pp. 487-568.
- Teo, Y. H., 2014. Water services industry reforms in Malaysia. *International Journal of Water Resources Development*, 30(1).
- TheSun, 2019. *Air Selangor records best ever non-revenue water average in 2019*, Kuala Lumpur: The Sun Daily.
- TheSun, 2019. *Govt to reduce non-revenue water to 31% by end of 11MP*, Kuantan: The Sun Daily.
- Thomas, N., Daniel, J. & Jay, N., 2013. Acoustic emission leak detection on a metal pipeline buried in sandy soil. *Journal of Pipeline Systems Engineering and Practice* , 4(3), pp. 149-1554.
- Thompson, M., Chapman, C. J., Howison, S. D. & Ockendon, J. R., 2001. *Noise generation by water pipe leaks*. Keele University, Staffordshire, England.: 40th European Study Group with Industry.
- Torre, D. et al., 2015. Inverse Problems: Theory and Application to Science and Engineering 2015. *Mathematical Problems in Engineering*.
- Vitovsky, J. P., Simpson, A. R. & Lambert, M. F., 2000. Leak detection and calibration using transients and genetic algorithms. *Journal of Water Resources Planning and Management*, 126(4), p. 262–265.

- Wang, L. J., Zhang, H. W. & Hui, J., 2012. A Leak Detection Method Based on EPANET and Genetic Algorithm in Water Distribution Systems. *Software Engineering and Knowledge Engineering: Theory and Practice*, p. 459–465.
- Wang, X., Simpson, A., Lambert, M. & Vitovsky, J., 2001. Leak detection in pipeline systems using hydraulic methods: a review. *Conference on Hydraulics in Civil Engineering, (23-30):391–400, November 2001, 23(30)*, pp. 391-400.
- WaterWorld, 2019. *Malaysian water utility cuts non-revenue water by one-third with remote monitoring*, West Regions U.S.: WaterWorld Magazine.
- Whittle, A. et al., 2010. *A testbed for continuous monitoring of the water distribution system in singapore*, Singapore: Water Distribution System Analysis (WSDA).
- WWA, 2018. *Enigma3m trial in Johor pinpoints 115 leaks*, Singapore: Water & Wastewater Asia.
- Zafar, A., 2016. Comparison of Dijkstra's Algorithm with other proposed algorithms. *International Academic Journal of Science and Engineering*, 3(7), pp. 53-66.
- Zahab, S. & Zayed, T., 2019. Leak detection in water distribution networks: an introductory overview. *Smart Water*, 4(5).
- Zangenehmada, Z., 2014. *Study of leak detection technologies in water distribution networks*. Halifax, CSCE.
- Zangenhmadar, Z. & Moselhi, O., 2014. Study of leak detection technologies in water distribution networks. *CSCE 2014 General Conference*.

## LIST OF PUBLICATIONS

**W. H. Png, C. H. Pua, M. L. Tham, 2020. Multiple Paths Leak Localisation System in Water Distribution Network. *SSIP 2020 conference proceeding.***

**W. H. Png, H. S. Lin, C. H. Pua, M. L. Tham, F. A. Rahman, 2021. An Adaptive Leak Localisation System based on a Multi-level Analytics Framework in Piping Network. *International Journal of Sensor Network (IJSNET)*, 36(3), pp. 146-158.**