HASAN, M.J., ARIFEEN, M., SOHAIB, M., ROHAN, A. and KANNAN, S. 2024. Enhancing gas pipeline monitoring with graph neural networks: a new approach for acoustic emission analysis under variable pressure conditions. In *Proceedings of the 20th International conference on condition monitoring and asset management 2024 (CM 2024), 18-20 June 2024, Oxford, UK*. Northampton: BINDT [online], (accepted). To be made available at: https://doi.org/10.1784/cm2024.4b3

Enhancing gas pipeline monitoring with graph neural networks: a new approach for acoustic emission analysis under variable pressure conditions.

HASAN, M.J., ARIFEEN, M., SOHAIB, M., ROHAN, A. and KANNAN, S.

2024

Enhancing Gas-Pipeline Monitoring with Graph Neural Networks: A New Approach for Acoustic Emission Analysis under Variable Pressure Conditions

Md Junayed Hasan[1*], Murshedul Arifeen[1], Muhammad Sohaib[2,3], Ali Rohan[1], Somasundar Kannan[4]

[1] National Subsea Centre, Robert Gordon Uni., Aberdeen, AB21 0BH, UK
[2] School of Computer Sci. and Tech., Zhejiang Normal Uni., Jinhua 321004, China
[3] Zhejiang Inst. of Photoelectronics and Zhejiang Inst. for Adv. Light Source, Zhejiang Normal Uni., Jinhua 321004, China.
[4] School of Engineering, Robert Gordon Uni., Aberdeen, AB10 7AQ, UK

Correspondence (*): Md Junayed Hasan, email: j.hasan@rgu.ac.uk

Abstract

Traditional machine learning (ML) and deep learning (DL)-based acoustic emission (AE) data-driven condition monitoring models face several reliability issues due to factors such as fluid pressure changes, flange vibrations, inconsistent leak lengths, and noise in AE signals, which vary with pipeline conditions. Additionally, the noise, and variable pressure conditions complicate the interpretation of sensor data, especially in multi-variate setups where understanding spatial relationships between sensors is challenging. In response, we have introduced Graph Convolutional Networks (GCNs) to overcome these challenges in AE-based pipeline monitoring for the first time. Our proposed method utilizes a publicly available pipeline monitoring dataset, named GPLA-12, which comprises AE signals to train and evaluate the GCN-based model. This innovative graph construction technique is designed to decipher and comprehend the subtleties in AE signals gathered under various pressure conditions from a multi-variate sensor setup. This approach can potentially establish a new standard in pipeline monitoring research and applications.

1 Introduction

Pipelines play a pivotal role in transporting gases and fluids, with affordability and safety being paramount concerns. Despite their importance, pipelines are prone to leaks caused by various factors such as corrosion, fatigue cracks, material defects, environmental discontinuities, and seismic activities [1, 2]. These leaks can lead to severe economic losses, safety hazards, and environmental damage [3]. Statistics reflect the gravity of these incidents: the European Gas Pipeline Incident Data Group (EGIG) recorded 1,366 incidents in Europe between 1970 and 2016, with a failure frequency of 31% per year per 1,000 km [4]. In the US, significant incidents have also been documented, including 652 pipeline incidents up to February 2020, resulting in 36 injuries, 88 fires, and over $333 million in property damages. Additionally, between 2019 and 2021, the US Pipeline and Hazardous Materials Safety Administration (PHMSA) reported 267 incidents, 11 injuries, 21 fires, and $264 million in property damages [4, 5]. Given these alarming statistics, early leak detection and prevention strategies are crucial for mitigating such consequences.

Several techniques are available for detecting pipeline leaks, including negative pressure wave analysis, accelerometer-based detection, magnetic flux leakage detection, and acoustic emission (AE) technology [6, 7]. AE inspection is particularly favored for its early leak detection capabilities, real-time response, ease of installation, and sensitivity. Nevertheless, AE monitoring is challenged by the complexity of the signals, which require sophisticated algorithms for denoising, feature extraction, and analysis. These algorithms, however, typically cater to offline data and struggle with noise interference, making accurate categorization of AE events difficult under variable conditions. AE signals are inherently nonlinear and nonstationary, influenced by fluid dynamics and stress wave mechanics, necessitating advanced processing techniques and a deep understanding of the underlying physical processes [8, 9].

Recent advancements in machine learning (ML) and deep learning (DL) have transformed the field of condition monitoring by facilitating the representation of complex features, crucial for analyzing the intricate datasets generated in pipeline monitoring scenarios [10, 11, 12]. However, traditional AE monitoring methods still face issues such as fluid pressure variations, flange vibrations, and inconsistent leak lengths, which compromise data reliability [13,14]. Furthermore, noise in AE signals complicates the interpretation of data from sensors distributed along pipelines, especially under varying pressure conditions and in multi-variate sensor setups. This complexity is exacerbated by the necessity to analyze data within the confines of Euclidean space, which inadequately represents the nonlinear relationships found in real-world scenarios.

To address these challenges, we propose the use of Graph Neural Networks (GNNs) [15,16,17], an approach yet to be fully exploited in AE-based pipeline monitoring. Our study introduces a novel methodology that employs GNNs alongside a publicly available pipeline AE dataset, called GPLA-12 [18]. This innovative graph construction method is designed to enhance the understanding of AE signal nuances under different pressure conditions from a multi-variate sensor setup. We incorporated statistical properties from multiple sensors to train the model, enabling the network to learn the spatial relationships between sensors while simultaneously understanding the temporal dynamics within the data from individual sensors. Our approach successfully achieved a 96.2% accuracy rate in detecting changes under variable pressure conditions.

The rest of the paper is organized as follows: Section 2 refers to the experimental set up and framework strategy, Section 3 discusses the experimental finding, and finally, Section 4 concludes the paper.

2 Experimental setup and framework strategy

Fig. 1 depicts the complete framework for the validation of our case study. The process begins with data collection by the only open access AE pipeline dataset, GPLA-12 [18], followed by calculating statistical properties, denoted as AE-prop (Fig. 2). With the AE-prop step, we convert the data to be suitable for use as the input node in a GNN. We considered multiple training models with different layer and hyperparameter strategy, incorporating stratified k-fold cross-validation (CV) to determine the optimal model. Once satisfactory performance is achieved, the optimal model is selected for evaluating performance on test data.

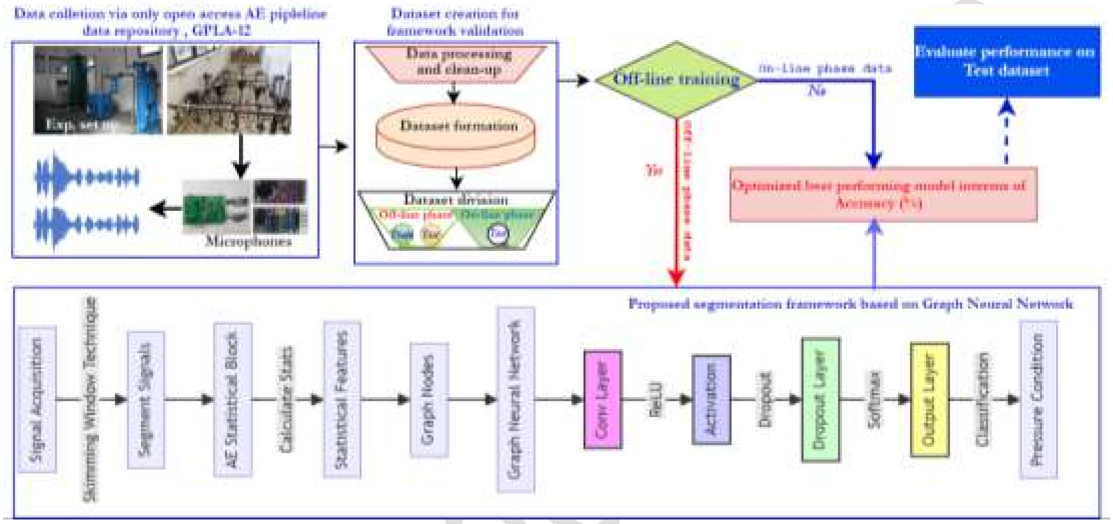

Fig. 1. Proposed method.

2.1 Dataset Description

In this study, we utilized the GPLA-12 dataset [18], which is publicly available and comprises multi-band sound sensors specifically designed to capture AE signals under various pipeline conditions within a simulated engineering environment. The dataset underwent a rigorous filtration process to eliminate contaminants such as moisture, dust, and oil. Data collection was performed using a custom-built signal transmission module, a single-chip microcomputer, and two audio sensors tailored to different frequency ranges. The pre-amplifier parameters are reported as follows: sensor 1 (MAX4466) is from 1000 Hz to 10 kHz, whereas sensor 2 (LM38) is from 0 Hz to 600 kHz. However, the microphones used are electret of which there are no references present in the dataset description. Despite the dataset's challenges, such as the lack of baseline signals, absence of leak size and location details, unstructured category arrangement, low feature density, and significant intra-class variance [12], we chose to employ it to validate our method because of its ability to address these limitations effectively. In this study, we analysed the dataset for 3 categories, each containing 228 AE samples with a discrete variable length of 1460 for 3 different gas pressure conditions. Table 1 provides category-specific information.

2.2 Model Training Strategy, and Evaluation

We have divided our training and testing strategy into two phases: online and offline. In the offline phase, we evaluate the model's performance using the training and validation sets. Once the optimal model is selected based on the highest accuracy, we use this model to assess performance on the test data during the online phase. According to our empirical study [19], we randomly allocated 80% of the data for training (offline phase) and 20% for testing (online phase). For the offline phase, we used stratified k-fold CV, allocating (1/k) % of the data for validation. To reserve approximately 30% of the data for validation, we set k to 5. Each fold involved 1600 epochs. While running each fold, we changed the hyperparameters, and layer architecture. Given the limited number of samples, constructing a generalized DL model was challenging. To address this, we employed a skimming window technique to further segment the signal and enhance feature density, as detailed in [20, 21, 22]. In the upcoming extension of this work, we are considering details of this aspect.

Table 1. Dataset details.

Pressure	Data amount				Total
		Offline phase (80%)		Online phase (20%)	
	Defined	Training	Validation*	Testing	
0.2 MP	Label 1	146	36	46	228
0.4 MP	Label 2	146	36	46	228
0.5 MP	Label 3	146	36	46	228

** Validation is done with Stratified k fold CV*

2.3 Graph Neural Network – Brief Description

In our study, as a part of GNN, we have considered Graph Convolution Neural (GCN) models work on our graph structured data. Therefore, time series data first needs to be transformed into graph structure to apply GCN. A graph G = (V, E, A), comprises a set of nodes V, | V | = K, a set of edges E, | E | = L, and an adjacency matrix A. The adjacency matrix A ∈ $R^{K\times K}$ contains the weights and edges among the graph nodes V. If there is an edge between node v_i ∈ V and v_j ∈ V, then they are neighbours i ≠ j, and the entry of A (i, j) in the adjacency matrix A depicts the weights of their edge. The weights of the edges can be calculated using various techniques, for example, Euclidean similarity, correlation matrix, or cosine similarity. GCN performs convolution in the vertex domain and updates a node's representation by recursively aggregating information from its neighbor nodes [15]. By representing neighborhood information as graph substructures, these substructures can be modelled using differentiable functions that recursively project different or identical substructures into other or the same feature spaces [16]. Messages, also known as information flow between neighbors and centre nodes [17]. Propagation rules of messages characterize network architecture based on message passing. The message passing rules among the nodes can be defined into two stages— the message passing stage and readout stage. Mathematically, we can represent these two stages as [17]:

$$M_i^{l+1} = \sum_{j\in N_i} Message(h_i^l, h_j^l, h_{ij}^l) \quad (1)$$

$$h_i^{l+1} = Update(h_i^l M_i^{l+1}) \tag{2}$$

Where, h_i, h_j, and h_{ij} represent two specific nodes and their connecting edge feature. Node i's neighboring nodes are represented by the set N_i. The neighboring information is transformed into a hidden representation by the Message (·) function and passed to the centre node. Update (·) function aggregates and updates the centre node's representation. The choice of Message (·) and Update (·) functions may vary.

3 Result analysis and discussion

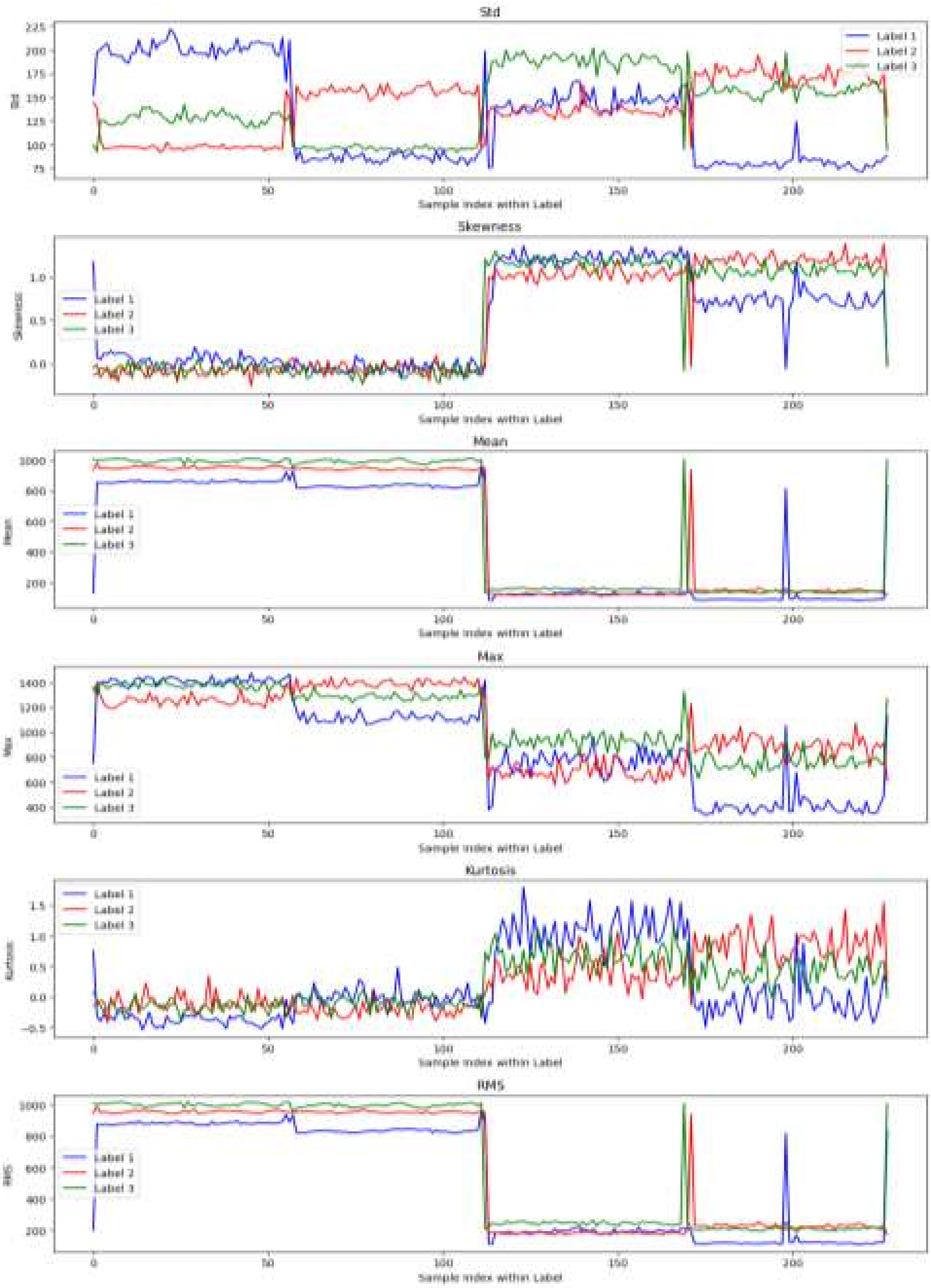

Fig. 2. AE-prop.

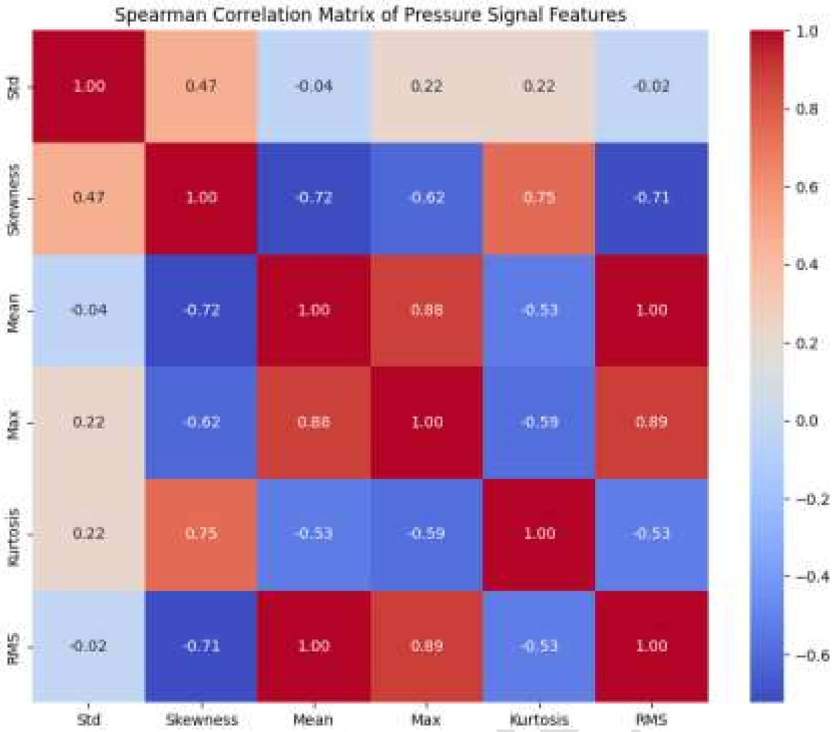

Fig. 3. AE-prop correlation.

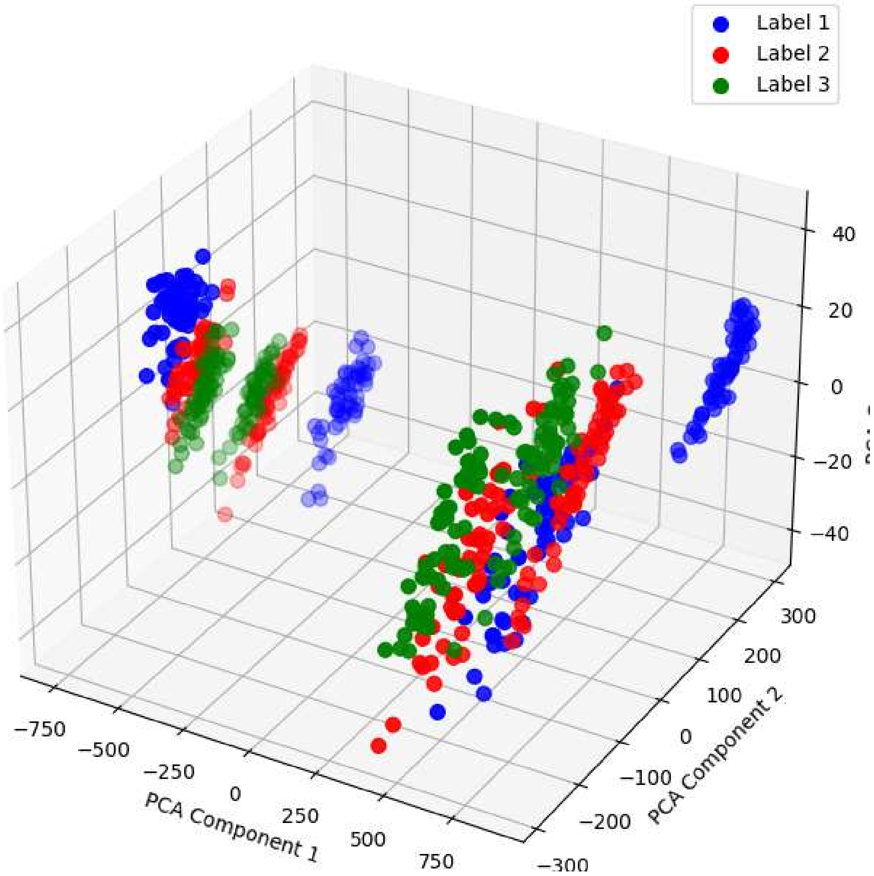

Fig. 4. 3D PCA for AE-prop.

Initially, we calculated the statistical properties of the data, which included the standard deviation (std), skewness, mean, maximum value, kurtosis, and root mean square (RMS) of the signal. The details of these analyses are presented in Fig. 2. Subsequently, we investigated the feature correlation, as shown in Fig. 3. However, we did not address the issue of multicollinearity by removing highly correlated features in this study, as it serves as a baseline for experimenting with GCN for AE based condition monitoring. In our forthcoming study, which extends this work, we will consider this aspect.

Following this, we employed Principal Component Analysis (PCA) to examine feature separation, depicted in Fig. 4. In both Fig. 2 and 4, it is evident that the feature separation for each label falls into two distinct ranges. For instance, the RMS values for label 1, as shown in Figure 2, drop after the 120th index, a pattern that is consistent across all labels. This observation suggests that the data was collected using two different sensors with varying pre-amplifier settings, as detailed in the dataset description. However, the objective of this study was not to consider these variances, but rather to classify the data as a whole under three different pressure conditions.

Once the AE-prop were calculated, we transformed the AE data into graph nodes using these properties. Subsequently, we commenced the training process in the offline phase. We employed 5-fold stratified CV across five variants of GCN. For each model variant, we reported the accuracy of the best-performing fold. Then, we selected the model variant that demonstrated the highest accuracy as the optimal one. Detailed information about the training process is presented in Table 2 and Fig. 5. Here, M3 and M5 performs very similarly, however, M5 is stable (Fig. 5) in terms of loss curve, and accuracy fluctuations. Therefore, as optimal model, we have chosen M5. However, on the online phase, the performance dropped, and we achieved 94.21% accuracy.

Table 2. Model performance details.

Phase	**Models**	**Specifications**	**Performance**	**Remarks**
Offline	**Variants**	**Layer details**	**Accuracy (%)**	
	M1	1 Convolution 1 Dropout	82.30	
	M2	2 Convolution 1 Dropout	83.91	
	M3	2 Convolution 2 Dropouts	95.13	
	M4	3 Convolution 1 Dropout	85.21	
	M5	**3 Convolution 2 Dropouts**	**95.62**	**Optimal**
Online	**Optimal**		**95.62**	

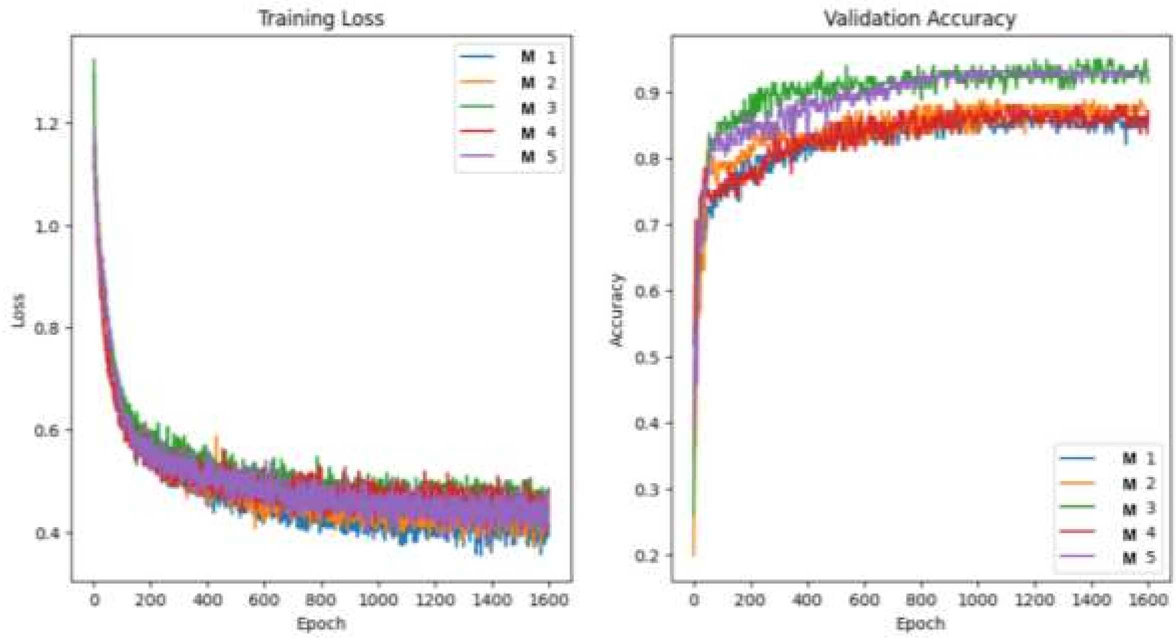

Fig. 5. Model performance (training loss and validation Accuracy – offline phase).

4 Conclusions

In this study, we have explored the use of GCN for AE driven pipeline condition monitoring. This case study was conducted using the only available online dataset for pipeline condition monitoring, the GPLA-12 dataset. A significant challenge was the lack of detailed information about this dataset, making it difficult to validate outcomes against sensor specifications. Additionally, during the design of this experiment, we did not address the multicollinearity of statistical features, AE-driven feature analysis, or the generation of synthetic data to increase the sample size. Omitting these considerations poses challenges in achieving generalized performance, as indicated by our online phase test accuracy of 95.62%. This study establishes the baseline for our ongoing case study. In the extended version of our work, we will address these issues and offer a detailed analysis. Nevertheless, the aim of this initial study is to demonstrate the potential of GCNs to handle multi-sensor setups and to explore temporal and spatial relationships while reporting performance metrics. Although the results are comparable to those of conventional ML/DL approaches, these considerations justify further exploration of GNNs.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT, and Quill Bot in order to refine the writing, in-terms of language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

1. Quy TB, Kim JM. Pipeline Leak Detection Using Acoustic Emission and State Estimate in Feature Space. *IEEE Trans Instrum Meas*. 2022;71:1-9.
2. Zhang B, Kang R, Zhu H, Yuan Q, Zhang Q, Zhou X. Numerical simulation and safety evaluation of multi-source leakage of buried product oil pipeline. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*. 2022;44(3):6737-6757.
3. Zhang Z, Zhang L, Fu M, Ozevin D, Yuan H. Study on leak localization for buried gas pipelines based on an acoustic method. *Tunnelling and Underground Space Technology*. 2022;120:104247.
4. European Gas Pipeline Incident Data Group Database. EGIG Report. Published 2020. Accessed February 2, 2023. https://www.egig.eu/reports
5. United States Pipeline and Hazardous Materials Safety Administration. US PHMSA. Accessed February 2, 2023. https://www.phmsa.dot.gov/
6. The American Society for Nondestructive testing. *Nondestructive Testing Handbook, Third Edition: Volume 6, Acoustic Emission Testing (AE).*; 2005. Accessed February 2, 2023. https://asnt.org/
7. Feng J, Li F, Lu S, Liu J, Ma D. Injurious or noninjurious defect identification from MFL images in pipeline inspection using convolutional neural network. *IEEE Trans Instrum Meas*. 2017;66(7):1883-1892.
8. Hu X, Zhang H, Ma D, Wang R. A tnGAN-based leak detection method for pipeline network considering incomplete sensor data. *IEEE Trans Instrum Meas*. 2020;70:1-10.
9. Rai A, Kim JM. A novel pipeline leak detection approach independent of prior failure information. *Measurement*. 2021;167:108284.
10. Banjara NK, Sasmal S, Voggu S. Machine learning supported acoustic emission technique for leakage detection in pipelines. *International Journal of Pressure Vessels and Piping*. 2020;188:104243. doi:https://doi.org/10.1016/j.ijpvp.2020.104243
11. Mujtaba SM, Lemma TA, Vandrangi SK. Gas pipeline safety management system based on neural network. *Process Safety Progress*. 2022;41:S59-S67.
12. Liao H, Zhu W, Zhang B, et al. Application of Natural Gas Pipeline Leakage Detection Based on Improved DRSN-CW. In: *2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT)*. IEEE; 2021:514-518. Accessed May 4, 2023. https://ieeexplore.ieee.org/document/9696455
13. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *European Conference on Computer Vision*. Springer; 2016:630-645. doi:https://doi.org/10.1007/978-3-319-46493-0_38
14. Zhao M, Zhong S, Fu X, Tang B, Dong S, Pecht M. Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis. *IEEE Transactions on Industrial Electronics*. 2020;68(3):2587-2597.
15. Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F. Graph convolutional networks for computational drug development and discovery. Briefings in bioinformatics. 2020 May 21;21(3):919-35.
16. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826. 2018 Oct 1.
17. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. InInternational conference on machine learning 2017 Jul 17 (pp. 1263-1272). PMLR.

18. Li J, Yao L. GPLA-12: An Acoustic Signal Dataset of Gas Pipeline Leakage. *arXiv preprint arXiv:210610277*. Published online 2021.
19. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *International Journal of Intelligent Technologies & Applied Statistics*. 2018;11(2).
20. Hasan MJ, Islam MM, Kim JM. Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions. Measurement. 2019 May 1;138:620-31.
21. Hasan MJ, Kim JM. Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning. Applied Sciences. 2018 Nov 22;8(12):2357.
22. Hasan, Md Junayed, Jia Uddin, and Subroto Nag Pinku. "A novel modified SFTA approach for feature extraction." In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, pp. 1-5. IEEE, 2016.