

K-Means Clustering for Seismic Signal Classification using DAS Geophone Integrated Signals

Long Muhammad Haziq Long Hassan¹, Nurfarhana Mohamad Sapiee¹, Seri Mastura Mustaza² and Norhana Arsad^{1,*}

¹Photonic Technology Research Group, Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia UKM, 43600 Bangi, Malaysia

²Spectrum Technology (Spectech) Research Group, Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia UKM, 43600 Bangi, Malaysia

*Corresponding author: noa@ukm.edu.my

Keywords

*Distributed Acoustic Sensing (DAS)
Geophone
Data Fusion
Seismic*

Article History

Received: 18 May 2025

Accepted: 5 June 2025

Published: 30 October 2025

Abstract

Distributed Acoustic Sensing (DAS) and geophone technologies are implemented for seismic monitoring with limited comparative analysis between these systems particularly in integrating data from heterogeneous sensors. One of raising in utilization of advanced machine learning for seismic data classification. Data fusion is promising solution, which capable of combining information from multiple sources to improve accuracy and robustness of seismic interpretation. Hence, in this study focus on integrating DAS and geophone data using data fusion by utilizing unsupervised machine learning with K-Means clustering to classify seismic signals. Seven clusters are used in this study which used PCA and t-SNE scatter plot for the K-Mean clustering. Results show that the proposed methodology demonstrates effectiveness in distinguishing between common and anomalous seismic events. Visualization and cluster analysis confirm the interpretability and practical value of this framework. From the findings, this study offers promising basis for anomaly detection and routine monitoring system with future works aimed at improving clustering techniques.

1. Introduction

DAS and geophone are widely used in earthquake monitoring systems to detect and analyze seismic events. DAS utilizes Rayleigh backscattering in optical fibers to capture changes in optical phase and convert it into seismic signals, while geophone directly detects ground motions using electromagnetic induction system [1], [2]. DAS are beneficial in term of easy installation and spatial coverage while geophone has high sensitivity and capture less frequency noise [3].

Even though both DAS and geophones have been used widely as independent instruments for earthquake monitoring and detection system, very little study have reported on the integration between DAS and geophone for earthquake monitoring system. Hence, this paper is to study the integration between DAS and geophone by using data fusion techniques for seismic classification and apply and validate machine learning algorithms for seismic events classification. The combination of these sensors helps to overcome the limitation of using a single DAS or a geophone, especially in weak or overlying signal detection. This study is fundamental for optimizing more consistent and less expensive, scalable seismic monitoring systems developed with unsupervised machine learning.

2. Methodology

Figure 1 shows the flow chart on collected DAS and geophone data simultaneously in a seismic monitoring experiment. The main unit measurements in this study are in millivolts based on geophone readings. For the data fusion, DAS reading was converted to millivolts using a calibration factor based on empirical observations and domain expertise to ensure consistent data.

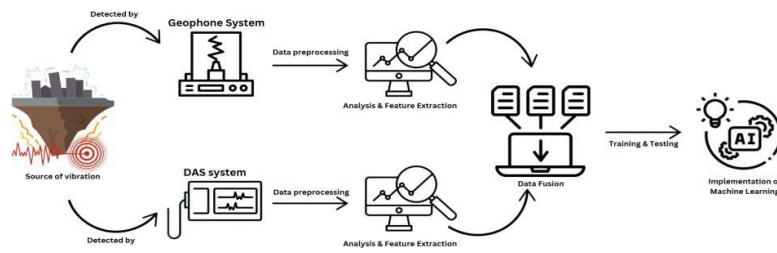


Figure 1. Flowchart of seismic monitoring experiment

For pattern classification, the geophone data sequence was divided into ten consecutive readings, with one-second interval. The maximum and mean absolute values that act as statistical features for each segment were calculated to represent the signal characteristics. All the feature values were then normalized to a 0-1 range using Min-Max Scaling. To classify the sequences based on these features, an unsupervised learning, K-Means clustering with seven clusters ($k = 7$) were applied to show different amplitude levels observed in the data [4]. The resulting clusters and their associated features were stored for further analysis.

3. Result and Discussion

The K-Mean Clustering successfully grouped the geophone readings into seven different clusters representing unique patterns in seismic data. Converting DAS readings to match geophone values for data fusion is important to create a unified dataset suitable for clustering. Maximum and mean absolute values were extracted as statistical features, normalized, and used to distinguish signal characteristics effectively. Figure 2 shows the PCA and t-SNE scatter plot for the K-Mean clustering (cluster = 7).

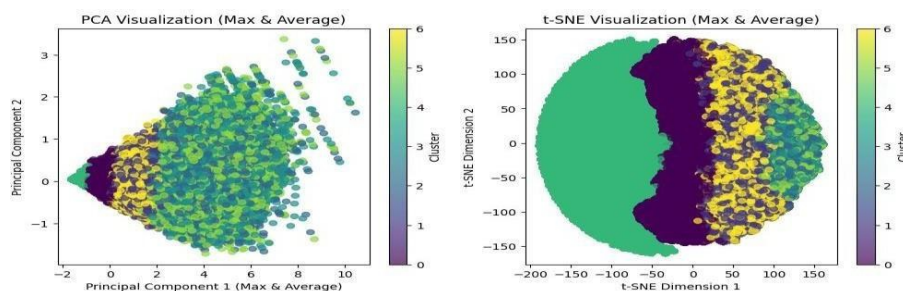


Figure 2. PCA/t-SNE scatter plot for 7 clusters

Based on the analysis of cluster distribution, it shows that certain amplitude levels are dominant than others. Clusters 0 and 4 were dominant with 247,200 and 244,455 samples respectively, showing recurring seismic conditions. Smaller clusters (2, 3, and 5) had only about 23,000 - 26,000 samples each, showing anomalous seismic events. The dominance of clusters 0 and 4 was confirmed using a bar chart of cluster frequencies, while a 2D scatter plots such as PCA and t-SNE showed that many clusters display discrete boundaries, while some of it overlap. Visualization using bar charts and time series plots shows how seismic reading evolved over time, which helps identify sudden spikes in amplitude, theoretically consistent with seismic events. Additionally, identifying amplitude associated with clusters was done by plotting a line graph, which reveals that certain clusters represented seismic activity with higher amplitude discrepancy while the others represented background noises.

The results show the proposed technique is effective for integrating DAS and geophone data and seismic signal pattern classifications using unsupervised machine learning. Most of the seismic activity detected during the monitoring from clusters 0 and 4 suggests that it was usual background noise or nature. However, the rare seismic events that were detected by small clusters of 2,3, and 5 could be representing seismic events such as microearthquake or man-made disturbances. Further analysis is needed to investigate these findings. Although the selected features effectively captured key patterns, cluster overlap suggests the need for more complex features. The current approach is effective and relevant. However, certain limitation persists. As this study relies on only two statistical features, all relevant aspects of seismic signals may not be captured. The assumption that the clusters are spherical and equally sized based on K-Mean Clustering cannot be applied to



all cases. Future work should include more diverse features, test using more flexible clustering algorithm to see the trends and do clusters validation based on known seismic events or data.

4. Conclusion

This study presents an effective and successful fusion between DAS and geophone data and seismic signal classification using unsupervised machine learning. By converting and scaling DAS readings to match geophone values, and using K-Means clustering with seven clusters, the methodology shows certain amplitude patterns of seismic readings responding to recurring daily activity while small numbers correspond to potentially anomalous events. Overall, the proposed methodology offers a practical framework for classifying seismic patterns automatically and detecting anomaly from DAS-geophone integrated datasets.

References

- [1] M. R. Fernández-Ruiz et al., "Distributed acoustic sensing for seismic activity monitoring," *APL Photonics*, vol. 5, no. 3, Mar. 2020, ISSN: 2378-0967. DOI: [10.1063/1.5139602](https://doi.org/10.1063/1.5139602)
- [2] K. Ma, J. Wu, Y. Ma, B. Xu, S. Qi, and X. Jiang, "An Effective Method for Improving Low-Frequency Response of Geophone," *Sensors*, vol. 23, no. 6, p. 3082, Mar. 2023, ISSN: 1424-8220. DOI: [10.3390/s23063082](https://doi.org/10.3390/s23063082)
- [3] S. R. Chandran, H. Mohamad, M. Y. M. Nasir, M. F. Ghazali, M. A. Abdullah, and V. Epin, "A Comparative Study of Seismic Characteristics Between Distributed Acoustic Sensing (DAS) and Geophones," in *Advances in Civil Engineering Materials*. Springer Nature Singapore, 2024, pp. 771–783, ISBN: 9789819707515. DOI: [10.1007/978-981-97-0751-5_66](https://doi.org/10.1007/978-981-97-0751-5_66)
- [4] H. Hartono, H. Anwar, R. Umam, and H. Takahashi, "An unsupervised machine learning algorithm approach using K-Means Clustering for optimizing Surface Wave Filtering in seismic reflection data," *Journal of Natural Sciences and Mathematics Research*, vol. 10, no. 1, pp. 114–127, Jul. 2024, ISSN: 2614-6487. DOI: [10.21580/jnsmr.v10i1.22399](https://doi.org/10.21580/jnsmr.v10i1.22399)