DOI: doi.org/10.21009/03.1401.FA14

Early Detection of Seismic Signal Anomalies Using Raspberry Pi 5 and Lightweight Machine Learning Models

Ahmad Kadarisman[1,a)], Imam Fachruddin[1)], Santoso Soekirno[1)], Hanif Andi Nugraha [2)], Benyamin Heryanto Rusanto[3)], Martarizal[1,b)]

[1] *Departemen Fisika, Univ. Indonesia, Depok, Jawa Barat 16424, Indonesia.*
[2] *Direktorat Instrumentasi dan Kalibrasi, BMKG, Jl. Angkasa I/2 Kemayoran - Jakarta, Indonesia.*
[3] *Sekolah Tinggi Meteorologi, Klimatologi dan Geofisika, Jl. Meteorologi No.5, Tangerang-Banten, Indonesia.*

Email: [a)] ahmad.kadarisman@ui.ac.id, [b)] martarizal@sci.ui.ac.id

Abstract

Data integrity is crucial for seismic monitoring systems, but is often compromised by anthropogenic or instrumental anomalies. This paper proposes a lightweight edge computing framework using Raspberry Pi 5 for real-time anomaly detection. MiniSEED data from the high-noise TOJI station were processed through segmentation, statistical or spectral feature extraction, and unsupervised models (isolation forest and autoencoder). The results show a detection latency of 78-113 ms with minimal resource consumption (<35% CPU, <200 MB RAM) and 82% correlation with ground-truth anomalies. This framework can be used on networked seismographs with limited resources such as those of the BMKG.

Keywords: Anomaly detection, Raspberry Pi 5, Isolation Forest, Autoencoder.

INTRODUCTION

Continuous monitoring of seismic activity within the Earth's crust is a fundamental component of governmental disaster mitigation strategies for earthquakes [1]. To this end, the Meteorology, Climatology and Geophysics Agency (BMKG) operates a vast network of seismographs to detect tectonic events at the earliest possible moment. Unfortunately, the accuracy of these data is frequently compromised by anomalous signals unrelated to natural seismic events [2].

These contaminant signals stem from two primary sources : anthropogenic disturbances (also known as cultural noise) from human activities such as traffic and industry, and instrumental instabilities such as electronic glitches or sensor resonance [3], [4]. Such disturbances can mimic the characteristics of weak earthquakes, leading to false alarms and diminishing the overall reliability of the monitoring system [2]. For this study, the TOJI station (Tomo-Sumedang) as shown in FIGURE 1 was selected as a case study due to its proximity to a major highway, which results in data heavily contaminated with anomalies.

Historically, data validation has typically been performed centrally, which is inefficient due to network latency issues, high bandwidth requirements, and the risk of single points of failure. Centralized data validation is an inefficient model because of network latency, high bandwidth requirements, and the risks associated with single points of failure [5]. Edge computing architectures are more promising because they move data processing computing units closer to the data source [6], which is expected to create low-latency noisy environmental seismic signals, thus allowing us to use more affordable devices to process seismic signals, such as the Raspberry Pi 5 which has specifically proven to be capable of analyzing local seismic signals. Based on this approach, we aim to design the most lightweight and automated detection framework by leveraging unsupervised machine learning methods to obtain and improve the quality of the processed data and mobile systems across the monitoring network [7].

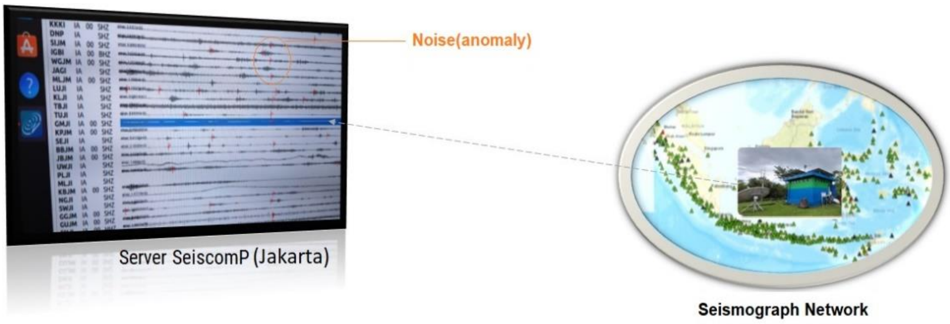

FIGURE 1. A diagram illustrating the seismic signal monitoring process within the BMKG seismograph network.

RESEARCH METHODOLOGY

System Architecture

The anomaly detection framework is built upon a decentralized architecture that places computational processing directly at the monitoring station [5]. At the core of this system is the Raspberry Pi 5, a single-board computer chosen for its cost-effectiveness, low power consumption, and sufficient computational power [7]. As illustrated in FIGURE 2, the Raspberry Pi is inserted between the seismograph's digitizer and the communication device. This strategic placement allows the device to process the seismic data stream in real time before it is transmitted to a central server [6]. This on-site approach offers substantial advantages, including ultra-low detection latency, reduced bandwidth needs, enhanced data privacy, and greater fault tolerance compared to centralized models [5], [8].

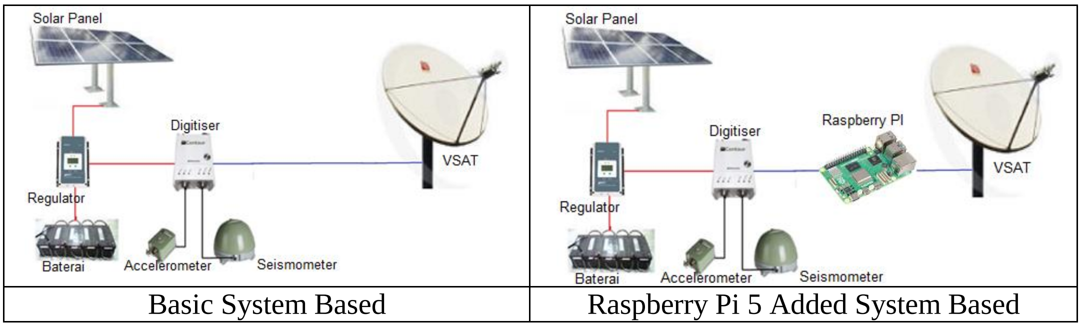

FIGURE 2. The proposed system architecture shows the integration of the Raspberry Pi 5 into the seismic station's data flow for on-site processing.

Data Acquisition and Pre-processing on Raspberry Pi 5

Seismic data is acquired from the digitizer in miniSEED format, the global standard for seismic data exchange as shown in FIGURE 3. The pre-processing pipeline is as follows:

- Acquisition: A Python-based framework using ObsPy and SeisComP's seedlink protocol [7] receives the real-time data stream from the digitizer.
- Segmentation: The continuous data are divided into time windows (e.g., 60 seconds). These segments are buffered for analysis, including triggering via a Short-Term Average/Long-Term Average (STA/LTA) algorithm [8] before being forwarded as a clean miniSEED stream [9].
- Identification and Filtering: Signal characteristics are identified using spectral analysis techniques such as fast Fourier transform (FFT) and power spectral density (PSD) [10] to

obtain a stable, noise-free spectral estimation. A bandpass filter is applied to remove irrelevant frequency components (e.g., below 0.1 Hz and above 20 Hz) [12].

- Processing: A hybrid approach combining time-frequency transforms and machine learning is used for noise reduction. The signal-to-noise ratio (SNR) is calculated as a key indicator of the resulting data quality [11].

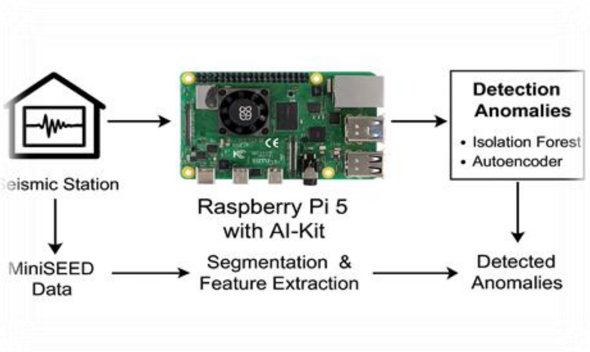

FIGURE 3. A flowchart of the data acquisition and pre-processing pipeline on the edge device.

Data Feature Extraction

For each data segment, a set of statistical and spectral features is extracted to create a representative vector for the machine learning models []:

- Statistical Features (Time Domain): Root Mean Square (RMS) to measure signal energy, Curtosis sensitive to sharp spikes fluctuation (glitch), and Skewness to measure asymmetricity amplitude distribution [10].
- Spectral Features (Frekuency Domain): The vector power spectral density (PSD) represents the power distribution across different frequencies set of values, and the high frequency ratio in high scale bandwidth (high frequency ratio) is crucially to distinguish earthquake waveform tremors that coming out from anthropogenic all noises [13].
- Zero crossing rate and peak-to-peak amplitude: To define any abrupt changes in the signal.

These features are compiled into a fixed-dimension feature vector that serves as the input for the detection models.

Anomaly Detection with Machine Learning

These features are compiled into a fixed-dimension feature vector that serves as the input for the detection models such as:

A. Isolation Forest

His algorithm is specifically designed for efficient outlier detection by constructing a collection of random decision trees (iTrees) [13], [15]. As an unsupervised method, it does not require labeled data for training, making it ideal for distinguishing between normal and anomalous signals [15]. Its Core Principle involves :

- Data Isolation: The algorithm creates multiple random trees, each of which is shaped by applying any random feature. Then, in a recursive way, each of the trees would separate data into a smaller subset with different feature interpretation (e.g., amplitude, frequency, duration, etc) and create a few split points into the feature range. These processes will last until all the data points are isolated or reach the deepest tree feature range.
- Isolating path, anomalous data have different attributes and solid clustered normal data commonly need fewer snippings or splitting which make a longer trail.
- An Anomaly Score, After a couple of isolated trees were built, the next algorithm will count the anomaly score of each data set. The standard of the scorization was taken by measuring the average of the isolation path lengths in the trees for each data point. A score greater than 0.5 typically indicates an anomaly [16], [17].

 p-ISSN: 2339-0654
e-ISSN: 2476-9398

- Anomalies, being "few and different," are easier to isolate and will thus have shorter average path lengths in the trees.

B. Lightweight Autoencoder

An autoencoder is a type of neural network trained to learn an efficient representation (encoding) of data and then reconstruct the original data (decoding) from that representation [14], [18]. The model is trained exclusively on normal seismic data to minimize the reconstruction error. When a new, unseen data point is processed, a high reconstruction error suggests that the data does not conform to the normal patterns learned by the model and is therefore flagged as an anomaly. The architecture is kept "lightweight" with fewer layers and neurons to ensure low-latency performance on the Raspberry Pi 5 [19]. There are a few steps of the encoding process in the following order :

- Seismicall Data Preprocessing, including segmentation process, normalization/ standardization and feature extractions.
- Dataset Generation, by collecting a big number of normal seismic data (non-anomalious dataset) that representing generalized seismic conditions.
- Lightweight Autoencoder Architecture Designing, Choosing the number of layers, neuron for each layer and creating generating function (e.g., ReLU : Rectified Linear Unit; MSE : Mean Squared Error; Optimizer).
- Encoder training, training the machine learning model using the seismically normalized data from the preprocessing phase, to distinguish the noise from the data.
- Real Data Anomaly Detection, Input the real seismic data to the trained autoencoder model.
- Thresholding, statistically determined from the reconstruction error distribution in trained normal data (e.g., in 95% or 99% percentages, or Isolation Forest in Construction error domain). For each data that exceed the threshold limit will be categorized as an anomaly [18].

RESULTS AND DISCUSSION

The system was tested using approximately 30 days of seismic data from the TOJI station, sampled at 40 Hz. The data were segmented into 3600 second windows for feature extraction such as RMS, curtosis, skewness, and spectral analysis by fast fourier transform (FFT) and power spectral density (PSD). The analysis results can be seen in FIGURE 4. Data analysis showed that the effectivity test system on the Raspberry Pi 5 works quite efficiently.

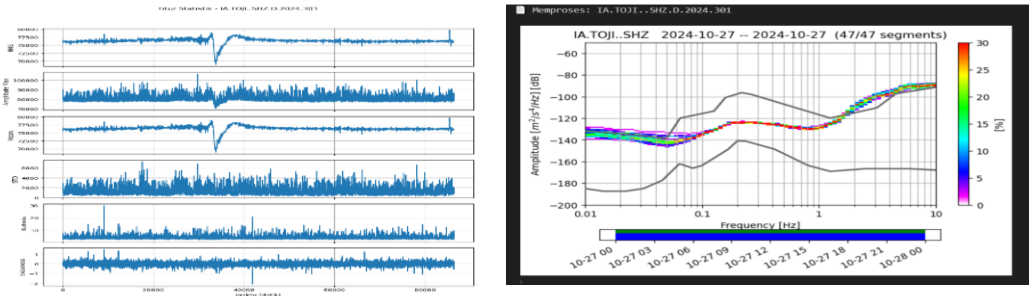

FIGURE 4. Example of the Feature Extraction Process and PSD Analysis from TOJI Station Data

From the TOJI Station Seismical data, there are found some implications and information, such as :

- Anomaly Detection Rate: The isolation forest model identified an average of 864 anomalies per day, as shown in FIGURE 5, while the autoencoder detected approximately ± 675 anomalies per day, as shown in FIGURE 6, respectedly to reconstruction parameters. The raw, noisy signal exhibited large, irregular amplitude fluctuations. PSD analysis confirmed that the anomalous signals had energy spread broadly across frequencies as shown in FIGURE 7, unlike natural seismic signals, which concentrate energy at lower frequencies [4].

p-ISSN: 2339-0654
e-ISSN: 2476-9398

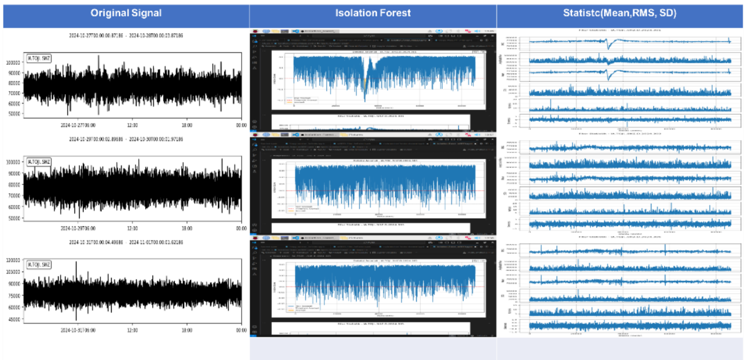

FIGURE 5. Results of anomaly detection using the isolation forest model.

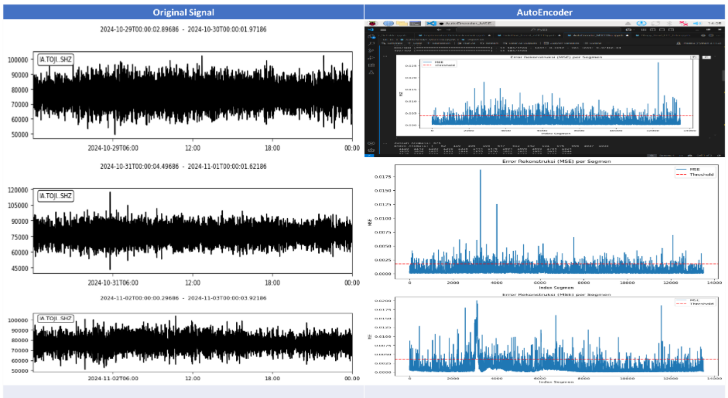

FIGURE 6. Results of anomaly detection using the autoencoder model.

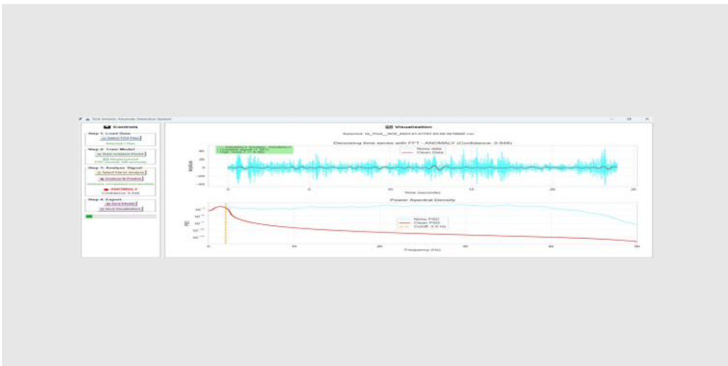

FIGURE 7. Comparison of a raw signal (top) versus a cleaned signal, with the corresponding power spectral density (PSD) plot (bottom) confirming the anomalous nature.

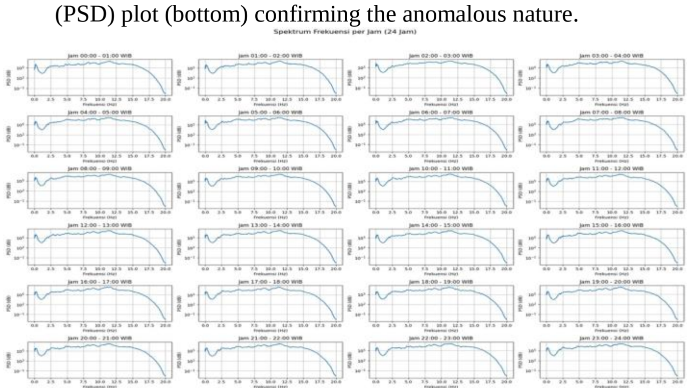

FIGURE 8. Seismic data frequency spectrogram during 24 hours. Transient energy surges (glitches) and continuous energy bands can be seen, suggesting instrument resonance and persistent anthropogenic disturbances.

 p-ISSN: 2339-0654
e-ISSN: 2476-9398

FIGURE 8 displays the spectrogram analysis from the one-day long data take. It shows some shady line patterns and energi fluctuation around the bandwidth frequency at 10-12 Hz, especially in day light [3]. These vertical lines indicate a momentary glitch ada pita, meanwhile the energy fluctuation showed us an instrumental resonance [20]. Moreover there are noise fluctuation patterns that are consistent with the human activity daily pattern (anthropogenic noise) [3]. Latensi Inference Latency: Inference time mean per segment is 78 ms (Isolation Forest) and 113 ms (Autoencoder) tested in Raspberry Pi 5.

- Resource Usage : CPU usage when the inference rate is below 35% and memory does not exceed 200 MB.
- Manual Log Correlation: Approximately 82% of the anomalies detected by the system corresponded with entries in a manual log of external activities, such as heavy traffic or local vibrations, validating the system accuracy.

TABLE 1. Performance Metrics of the Anomaly Detection Models on the Raspberry Pi 5.

Model	Avg. Anlomalies per Day	Avg. Inference Latency (ms)	Max. CPU Consumption (%)	Max. RAM Consumption (MB)
Isolation Forest	864	78	< 35%	< 200
Autoencoder	675	113	< 35%	< 200

Feature importance analysis from the Isolation Forest model (FIGURE 9) revealed that features related to frequency content and signal variability (high_freq_ratio, peak_count, var) were the most influential in detecting anomalies. This underscores the critical role of frequency-domain analysis over basic statistical metrics.

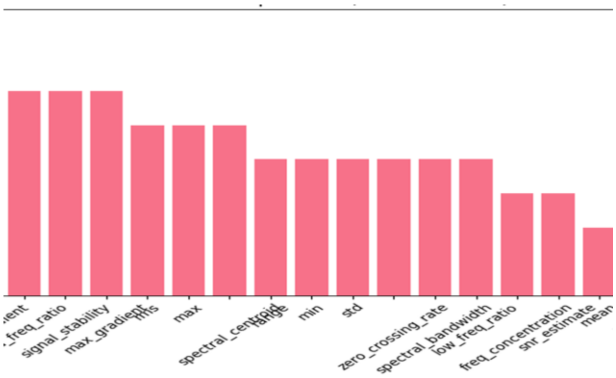

FIGURE 9. Feature importance ranking from the isolation forest model.

The experimental results identified several anomaly patterns while maintaining very low latency and minimal power consumption, making this framework feasible for implementation at other stations with limited infrastructure. The Isolation Forest model tends to be more sensitive to extreme outliers, such as sharp electronic noise [16], while the Autoencoder is more accurate in detecting subtle deformations in the signal waveform [18].

The system's success also depends on the selection of statistical and spectral features that fully represent anthropogenic noise. Although the system operates without a manual labeling mechanism, its minimal resource footprint (35% CPU, under 200 MB RAM) and high correspondence rate of 82% with manual activity logs underscore its significant potential for long-term station quality monitoring or in Early Warning Systems (EEW) [1][7].

A key challenge in real-time anomaly detection on edge devices such as the Raspberry Pi is finding a balance between the complexity of the features used and the limitations of the hardware's processing capabilities. In our experiments, we found that not all statistical and spectral features contribute proportionally to the detection performance [10], [13], as shown in TABLE 1.

For example, although the power spectral density (PSD) is considered a highly informative feature for seismic signals, its calculation using Welch's method in each time window imposes a significant computational burden (due to the fast fourier transform) on the Raspberry Pi. When PSD is combined with other complex features, such as spectral entropy or wavelet energy, the inference time increases significantly, up to 2.5 times, compared to using only basic features such as RMS, skewness, and kurtosis [12].

Another aspect often overlooked in research is model adaptation to dynamic environmental conditions, such as heavy rainfall, strong winds, or local disturbances such as traffic. This approach involves simply updating the basic statistical parameters of the features (e.g., mean and variance) without requiring a complete retraining of the entire model [2] [13].

CONCLUSION

This research successfully developed and validated an effective system for real-time, on-site anomaly detection in seismic data. The key conclusions are:

- High efficiency on the edge device : The system, tested on a Raspberry Pi 5 with data from the TOJI station, proved capable of detecting seismic signal disturbances in real-time with high efficiency. Its lightweight design makes it ideal for pre-processing and early detection within a larger seismic network.
- Unsupervised Learning Efficacy: The unsupervised approach, using an isolation forest and an autoencoder, effectively identified deviations from normal seismic patterns without the need for manually labeled datasets.
- Suitability for Resource-Constrained Environments: With its low cost, minimal power consumption (5 V), and small footprint, the system is perfectly suited for remote seismic stations that may rely on solar power and lack robust infrastructure.
- Scalability Potential: Given BMKG's extensive network of approximately 507 stations, this framework has significant potential for large-scale deployment. Future work should focus on inter-node network integration and enhancing the models' adaptability to long-term environmental changes.

ACKNOWLEDGEMENTS

The author would like to thank: Ayu Widowati, Marrissa Arlinkha, Rina Yuniarty, and Rini Anggraeni for their contributions to this research, as well as the Meteorology, Climatology, and Geophysics Agency (BMKG) through the Cooperation Agreement Letter PKS-012/UN2.F3.D/ PPM.00.02/2024.

REFERENCES

[1] U. Ashraf et al., "Application of Unconventional Seismic Attributes and Unsupervised Machine Learning for the Identification of Fault and Fracture Network," Applied Sciences, vol. 10, no. 11, p. 3864, Jun. 2020.

[2] J. Cai et al., "Research on U-Net Seismic Signal Denoising Combined with Residual Dense Blocks," Measurement Science and Technology, vol. 35, no. 5, p. 056104, May 2024.

[3] M. Carratù, E. De Benedictis, G. D'Elia, and M. D. Falco, "A Novel Methodology for Unsupervised Anomaly Detection in Industrial Electrical Systems," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1–12, 2023.

[4] P. P. Fastykovsky, S. V. Fastykovsky, and D. V. Fastykovsky, "Autonomous Compact Seismic Device for Detection and Recognition of Moving Person and Vehicles," in Proc. 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, 2020, pp. 969–972.

[5] F. Finazzi, "Earthquake Detection at the Edge: IoT Crowdsensing Network," Future Internet, vol. 14, no. 5, p. 143, May 2022.

[6] J. Gu, S. Yu, and D. P. G. D. P. Trad, "Unsupervised Deep Learning for Off-the-Grid Seismic Data Regularization and Denoising," Geophysics, vol. 89, no. 3, pp. 1–15, 2024.
[7] S. Han, D. Kim, and H. Song, "AI-Based Anomaly Detection Integrating Autoencoders and Isolation Forests," Medium, Jan. 23, 2024. [Online]. Available: https://medium.com/@han.s.ml/ai-based-anomaly-detection-integrating-autoencoders-and-isolation-forests-e9b995f214b5
[8] B. Housner, "AI on the Edge: IoT and Edge Computing Redefine Data Architectures," Database Trends and Applications (DBTA), Jan. 19, 2024. [Online]. Available: https://www.dbta.com/Editorial/Trends-and-Applications/AI-on-the-Edge-IoT-and-Edge-Computing-Redefine-Data-Architectures-162310.aspx
[9] A. Carvalho, C. Machado, and F. Moraes, "Raspberry Pi Performance Analysis in Real-Time Applications with the RT-Preempt Patch," in Proc. 2019 Latin American Robotics Symposium (LARS), Rio Grande, Brazil, 2019, pp. 162–167.
[10] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proc. 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413–422.
[11] H. Ma, Y. Liu, and H. Chen, "Noise Reduction for Desert Seismic Data Using Spectral Kurtosis Adaptive Bandpass Filter," Acta Geophysica, vol. 67, pp. 123–131, 2019.
[12] D. E. McNamara and R. P. Buland, "Ambient Noise Levels in the Continental United States," Bulletin of the Seismological Society of America, vol. 94, no. 4, pp. 1517–1527, Aug. 2004.
[13] S. M. Mousavi and G. C. Beroza, "A Machine-Learning Approach for Seismic Analysis of Industrial Facilities," Journal of Seismology, vol. 26, pp. 1–15, 2022.
[14] N. Nakata, L. Gualtieri, and A. Fichtner, Seismic Ambient Noise. Cambridge, UK: Cambridge University Press, 2019.
[15] A. T. Ringler et al., "A Review of Timing Accuracy across the Global Seismographic Network," Seismological Research Letters, vol. 93, no. 1, pp. 317–331, Jan. 2022.
[16] M. Sakurada and T. Yairi, "Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction," in Proc. MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, 2014, pp. 4–11.
[17] L. Seydoux, R. Balestriero, and P. Poli, "Clustering of Continuous Seismic Data Using Deep Scattering Networks," Geophysical Journal International, vol. 223, no. 3, pp. 1733–1748, Dec. 2020.
[18] A. Tolea, A. Tiganescu, and C. Ionescu, "Analysis of Ambient Seismic Noise at Bucovina Seismic Array (BURAR), Romania," Romanian Reports in Physics, vol. 75, p. 705, 2023.
[19] P. H. Tran, C. Heuchenne, and S. Thomassey, "An Anomaly Detection Approach Based on the Combination of LSTM Autoencoder and Isolation Forest for Multivariate Time Series Data," in Developments of Artificial Intelligence Technologies in Computation and Robotics. World Scientific, 2020, pp. 485–492.
[20] L. Utkin, S. Zatenko, and A. Meldo, "Improved Anomaly Detection by Using the Attention-Based Isolation Forest," Algorithms, vol. 16, no. 1, p. 19, Jan. 2023.