A Unified Framework for Replicability of Anomaly Detection in Multiple Time Series: Enhancing GANF and RanSynCoders Pipelines with MFCC Feature Extraction for Acoustic Data

Cristofer Silva
Universidade Católica de Pernambuco
Recife, Brazil
cgss@ecomp.poli.br

Matheus Coelho
Universidade Católica de Pernambuco
Recife, Brazil
matheus.2020131313@unicap.br

Umberto Barros
Universidade de Pernambuco
Recife, Brazil
utbf@ecomp.poli.br

Andrea Maria N C Ribeiro
Universidade Federal de Pernambuco
Recife, Brazil
andrea.marianogueira@ufpe.br

Rodrigo de Paula Monteiro
Universidade de Pernambuco
Recife, Brazil
rodrigo.monteiro@poli.br

Diego Pinheiro
Universidade de Pernambuco
Recife, Brazil
dmpfs@ecomp.poli.br

***Abstract*—Anomaly detection, a ubiquitous pattern recognition methodology for identifying atypical data, has been applied across a myriad of domains. In the industry, it has been used to identify malfunctions in industrial machinery in advance and suggest predictive maintenance based on multiple time series of sensor data, including temperature and pressure. With the availability of high-quality acoustic data, our understanding of acoustic scenes involving industrial machines has enhanced. However, the state-of-the-art approaches in anomaly detection, involve increasingly complex data processing pipelines, making replication with acoustic data not fully understood. Our proposal are three-fold: (i) we developed MTSA (Multiple Time Series Analysis), a unified framework for replicability of state-of-the-art anomaly detection approaches; (ii) we implemented Hitachi, RANSynCoders, and GANF on MTSA—three state-of-the-art anomaly detection approaches— improving RANSynCoders and GANF pipelines with a feature extraction based on Mel-Frequency Cepstral Coefficients (MFCC); and (iii) we conducted a comparative analysis of these approaches on MTSA for the anomaly detection with acoustic data from valves, pumps, fans, and slide rails using the open dataset for malfunctioning industrial machine investigation and inspection (MIMII). Overall, our results, measured by AUC-ROC, suggest RANSynCoders and GANF can be improved to handle acoustic data effectively, since these models archived solid results such as 0.94 (95% CI, 0.92-0.97) and 0.79 (95%, 0.75-0.82) respectively. As a unified framework, MTSA can decompose complex pipelines into simple data processing building blocks, facilitating the replication of existing approaches and the development of novel models.**

***Index Terms*—anomaly detection, multiple time series, acoustic data, replicability, MFCC**

I. Introduction

Anomaly detection involves identifying patterns in data that deviate from expected norms within a given context and labeling such data as abnormal [1]. It has gained increasingly relevance given its wide range of applications, specially in the industry, where it has been used to detect defects in industrial machinery [2]. In the industrial sector, malfunctioning machinery contributes to significant economic and safety burdens [3], which can be mitigated through predictive maintenance using machine learning models that learn from sensor data, including vibration, temperature and pressure [4].

The high-quality availability of acoustic data on industrial machinery has enhanced our understanding of anomaly detection but is still overlooked [5]. The study of acoustic events and scene analysis has created the yearly workshop on Detection and Classification of Acoustic Scenes and Events (DCASE). The public dataset for malfunctioning industrial machine investigation and inspection (MIMII) contains acoustic data collected from industrial machinery such as valves, pumps, fans, and slide rails operating under normal and abnormal conditions in a real factory environment [5].

The state-of-the-art approaches in anomaly detection, which handle multiple time series, often have complex data processing pipelines, bringing challenges to training and evaluate them in acoustic data [6]. Additionally, due to different code styles and project architectures, there is difficulties to use and replicate state-of-the-art models. The lack of a solid structure to adhere machine learning models jeopardizes the replicability in anomaly detection research and the exploitation of these model in complex data types—hindering the development of novel methods and the comparative analysis among existing approaches.

The contribution of this paper is three-fold. First, we implemented *MTSA* (Multiple Time Series Analysis), a unified framework facilitating the replicability of state-of-the-art anomaly detection approaches by decomposing complex

979-8-3503-7457-5/24/$31.00 ©2024 IEEE

pipelines into simple data processing building blocks, which can be easily reused and recombined to create novel pipelines. Second, using MTSA, we implemented Hitachi [5], RANSynCoders [7], and GANF [8]—three state-of-the-art anomaly detection approaches—improving the RANSynCoders and GANF pipelines with a feature extraction based on Mel-Frequency Cepstral Coefficients (MFCC), features commonly used in sound-related applications. Lastly, we conducted a comparative analysis of these approaches implemented on MTSA for the anomaly detection with acoustic data from valves, pumps, fans, and slide rails using the open dataset for malfunctioning industrial machine investigation and inspection (MIMII).

Our results suggest that RANSynCoders and GANF can be improved by incorporating an MFCC feature extraction at the beginning of their pipelines to effectively handle acoustic data. Overall, all models achieve comparable AUC-ROC results in detecting anomalies across the four machine types. For example, the GANF and Hittachi models had similar results for the slide rail machine, with 0.94 (95% CI, 0.92-0.97) and 0.99 (95% CI, 0.99-0.99), respectively, and RANSynCoders appears to be superior to all models for the pump machine with 0.79 (95% CI, 0.75-0.79). Additionally, parametric analysis within each model demonstrates the necessity of hyper-parameter tuning for each machine type.

The framework MTSA provide a solid environment to exploit models and helps to mitigate the replicability crises in anomaly detection research, facilitating the comparative analysis among existing approaches and the development of novel models.

The remainder of this paper is structured as follows: the Section II presents a review of models, Section III argues about methodology used in this paper, Section IV exhibit the background of the experiments, Section V presents the quantitative and qualitative results and Section VI binds others sections and concludes the paper.

II. Related Works

The related works we present a classic benchmarking as well as two state-of-the-art **??** II-C

In the related works, we present Hitachi (Subsection II-A), a classical benchmark in acoustic data. Besides, we present two cutting-edge approaches GANF (Subsection II-B)

A. Hitachi

Hitachi is an autoencoder model for unsupervised anomaly detection that considers a log-Mel spectrogram as an input feature [5]. Its encoder network architecture is $FC(Input, 64, ReLU)$, $FC(64, 64, ReLU)$, and $FC(64, 8, ReLU)$. Its decoder network is $FC(8, 64, ReLU)$, $FC(64, 64, ReLU)$, and $FC(64, Output, None)$, in which $FC(a, b, f)$ means a fully connected layer with a input neurons, b output neurons, and activation function f.

B. GANF

Graph-Augmented Normalizing Flows (GANF) [8] is a approach for anomaly detection of multiple time series using a graph neural network (GNN). Its inputs includes a Bayesian network—a directed acyclic graph modeling the dependency between time series—and a LSTM Encoders [9] to learn the temporal dependence of each series (see Figure 1). After calculating the dependence between the series, learning the temporal dependence, probability density is estimated using Normalizing Flow [10]. A low probability density indicates abnormality.

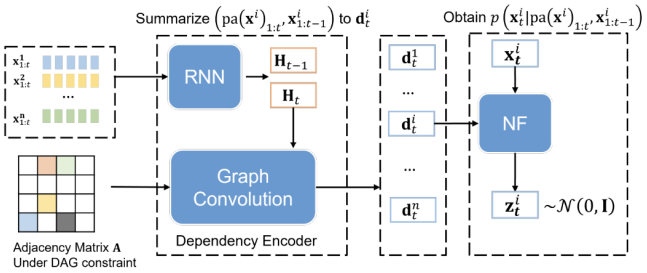

Fig. 1. Architeture of the GANF model obtained from [8]

C. RANSynCoders

RANSynCoders [7] is an approach for anomaly detection in multiple time series using a pipeline with the following sequential data processing steps (see Figure 2). The initial step involves dimensionality reduction using an autoencoder. It then estimates the latent spectral density using the Fast Fourier Transform (FFT), identifying the dominant frequencies in the time series, which are used as prior information to create a synchronized representation. The synchronization step transforms asynchronous time series into a unified representation. After synchronization, bootstrapping is used to reconstruct the upper and lower bounds of the full multivariate from each bootstrapped encoder. After training, anomaly inference is carried out using majority voting.

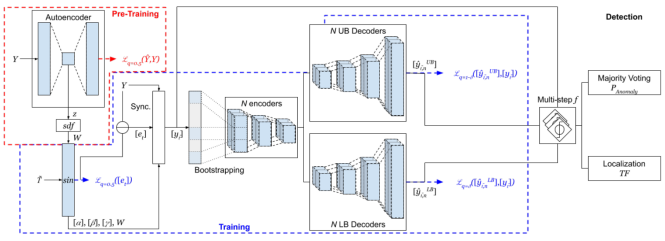

Fig. 2. Architeture of the RanSynCoders model obtained from [7].

III. Methods

The methodology of this work comprises the Acoustic Dataset from Industrial Machinery in Subsection III-A, the Mel-Frequency Cepstral Coefficients III-B, and Training and Testing III-C. All of the code is available on GitHub [11].

A. Acoustic Dataset from Industrial Machinery

The malfunctioning industrial machine investigation and inspection (MIMII) open dataset provides sounds from industrial machinery in both normal and abnormal operation. These sounds were collected using a TAMAGO-03 microphone with audio signals sampled at 16kHz. This microphone model is an array of eight distinct microphones in a circular shape. The existence of eight channels provides the opportunity to evaluate single-channel and multi-channel. The TAMAGO-03 was kept at a distance of 50 cm from the machines, 10 cm in the case of valves, and 10-seconds sound segments were recorded. The data from the MIMII dataset that we used in our study is showed in Table I.

TABLE I
ACOUSTIC DATA OF FOUR MACHINE TYPES UNDER NORMAL AND ABNORMAL OPERATING CONDITIONS FROM MIMII DATASET.

Machine Type	Model ID	Normal condition	Anomalous condition
Valve	00	991	119
Pump	00	1006	143
Fan	00	1011	407
Slide	00	1068	356
Total		4.076	1.025

B. Mel-Frequency Cepstral Coefficients (MFCC)

The Mel-Frequency Cepstral Coefficients (MFCC) [12] will be incorporated in the acoustic data preprocessing step. This feature extraction from the acoustic wave transforms the original signal into a smaller set of coefficients that capture the most important features of the sound and may facilitate easier processing by a computer [12]. This method extracts features from acoustic data using the Discrete Fourier Transform to obtain data in the frequency domain, then approximates it using the Mel filter for frequencies on the Mel scale. In this way, MFCC helps to reduce the dimensionality of the data, focusing on relevant spectral features of the audio signal even when the signal is contaminated by background noise. Given that Hitachi already considers a log-Mel spectrogram as an input feature, we only improved RANSyncoders and GANF by incorporating an MFCC acoustic data preprocessing step at the beginning their pipelines.

C. Training and Testing

The training and testing were conducted according to widely adopted unsupervised anomaly detection practices [5]. All the anomalous segments were reserved as the test dataset and an equal number of normal segments were randomly selected and also reserved for testing. The remaining normal segments were used as the training dataset. Using this training dataset, consisting only of normal segments, different instances of the models were trained and tested (see Figure 3). The training and testing were conducted using k-fold cross-validation, where this strategy segments the training data in k folds, and selects $k-1$ folds for training the models. After training, we validated the models using the same test data for each instance model.

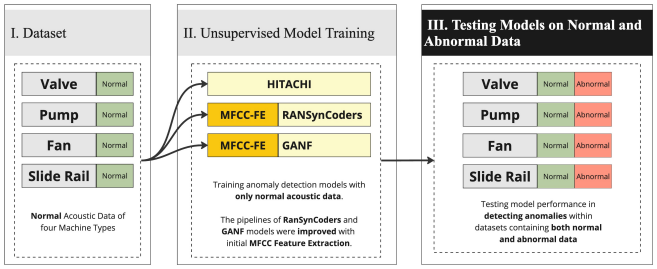

Fig. 3. Model training and testing setup.

IV. Experimental Setup

A. Synthetic Acoustic Data

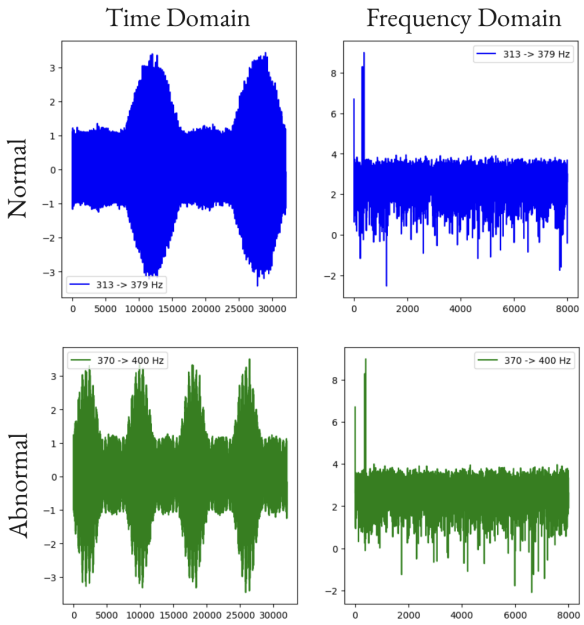

Fig. 4. Synthetic acoustic data on time (left) and frequency (right) domains for normal (top) and abnormal (bottom) data.

We have generated normal and abnormal synthetic acoustic data by controlling the wave frequency and the underlying noise level. This allows us to test the model with this data and better understand its behavior in different scenarios, such as high or low frequency sounds as well as sounds with different noise levels. In addition, it is possible to analyze which set of parameters performs best in these varying frequency and noise scenarios. The synthetic sound data is divided into normal and abnormal, mimicking the structure of the MIMII data set, where 100 are considered normal and 50 are considered abnormal. To generate the synthetic sounds, we created two

sinusoidal functions with different frequencies, 313 Hz and 379 Hz, and we add noise by Gaussian distribution with mean equals to 0 and standard deviation 0.1 before adding them together. We then repeated the process with two new sinusoidal functions with frequencies of 370 Hz and 400 Hz, also summing them after adding the same noise (Figure 4). We generated a total of 150 .wav files.

B. Models Hyper-Parameter Details

The optimization method for all models is the Adam. Custom hyper-parameters are described in Table II. Otherwise, standard hyper-parameters were used.

TABLE II
HYPER-PARAMETER MODELS SETUP

Model	Hyper-parameter	Value
Hitachi	Epochs	50
	Mels	64
	Frames	5
	Fast Fourier Transform	1024
	Hop Length	512
	Power	2
RANSynCoders	Epochs	10
	N Estimators	5
	Max Features	5
	Synchronize	True
	Min Periods	3
	Max Freqs	5
	Min Dist	60
GANF	Epochs	20
	Seed	10
	Hidden Size	32
	Weight Decay	5×10^{-4}
	Max Interaction	2
	Normalizing Flow Strategy	Masked Auto Regressive Flow

C. Statistical Analysis

The metric used to evaluate the trained models was the Area Under the Receiver Operating Characteristic Curve (ROC AUC). Confidence intervals were be constructed from the re-sampling with repetition using $1,000$ samples of the values obtained considering 5-fold cross validation. The use of statistical resampling methods makes it possible to generate estimates by producing additional samples based on an initial dataset. The bootstrap method, in particular, consists of generating random samples from the original sample. By applying the bootstrapping technique to the test results of the models (5 ROC AUC values), we can calculate the standard error and establish 95% for the confidence intervals.

V. RESULTS

The experimental setup was carried out to evaluate the performance of the Hitachi, RANSynCoders and GANF models in detecting anomalies in acoustic data.

A. Synthetic Acoustic Data

To evaluate the Hitachi, RANSynCoders and GANF models and help us to understand the results obtained in the real experiment described in section V-B, we conducted an experiment in a controlled environment using the synthetic acoustic data presented in section IV-A. The results obtained were satisfactory: all approaches performed well when trained and tested with the synthetic acoustic data, achieving a consistent AUC-ROC around of 1.0. These results indicate that the models can handle acoustic data and that they are able to operate efficiently with this type of information. In addition, the results obtained from this experiment reinforce the validity of the results obtained from real sound data (MIMII dataset).

B. Comparative Analysis

By comparing the best results obtained by the Hitachi, RANSynCoders and GANF models in different machines in MIMII dataset (see Table III), models are highly comparable given the overlap of CIs, with RANSynCoders being superior for pumps, achieving an AUC-ROC of .79 (.75 to .82). Additionally, anomaly detection appears to be easier for some machine types and more challenging for others. For instance, fans seem to be more challenging than slide rails.

TABLE III
COMPARATIVE ANALYSIS OF HITACHI, RANSYNCODERS, AND GANF ON VALVES, PUMPS, FANS AND SLIDE RAILS FROM THE MIMII DATASET. BOOTSTRAP 95% CONFIDENCE INTERVALS (CI) FOR THE AUC-ROC.

Model	Machine Type ID-00	AUC-ROC 95% CI
Hitachi	Valve	.56 (.45, .68)
	Pump	.77 (.75, .79)
	Fan	.60 (.59, .61)
	Slide rail	.99 (.99, .99)
RANSynCoders	Valve	.53 (.50, .55)
	Pump	.79 (.75, .82)
	Fan	.64 (.59, .69)
	Slide rail	.70 (.67, .73)
GANF	Valve	.58 (.48, .67)
	Pump	.73 (.68, .77)
	Fan	.54 (.52, .57)
	Slide rail	.94 (.92, .97)

C. Parameter Analysis

In this section, the results of the analysis of the Hitachi, RANSynCoders and GANF models hyper-parameters will be presented, including the variation of batch size and learning rate (see Table IV and Figure 5). By investigating how these hyper-parameters influence the model's performance, it is possible to identify optimal configurations that maximize performance in anomaly detection, contributing to the optimization and continuous refinement of the models.

Fig. 5. Comparative analysis of Hitachi, RANSynCoders and GANF, measured by bootstrap 95% confidence intervals (CI) for the AUC-ROC, with different hyper-parameter values (batch size and learning rate) on valves, pumps, fans and slide rails from the MIMII dataset.

TABLE IV
PARAMETER ANALYSIS OF BATCH SIZE AND LEARNING RATE FOR HITACHI, RANSYNCODERS, AND GANF, MEASURED BY BOOTSTRAP 95% CONFIDENCE INTERVALS (CI) FOR THE AUC-ROC, ON VALVES, PUMPS, FANS AND SLIDE RAILS FROM THE MIMII DATASET.

			AUC-ROC 95% CI			
Model	Lerning Rate	Batch Size	Pump	Fan	Valve	Slide Rail
Hitachi	1×10^{-6}	32	.81 (.79, .83)	.59 (.59, .59)	.30 (.29, .32)	.96 (.95, .97)
		90	.86 (.86, .86)	.57 (.57, .57)	.27 (.26, .27)	.98 (.98, .98)
		128	.86 (.85, .86)	.57 (.57, .57)	.28 (.27, .29)	.98 (.98, .98)
		360	.86 (.86, .87)	.57 (.55, .58)	.30 (.30, .30)	.98 (.98, .99)
		512	.86 (.85, .88)	.51 (.46, .56)	.30 (.30, .31)	.64 (.40, .89)
	1×10^{-3}	32	.75 (.73, .76)	.63 (.60, .66)	.74 (.71, .77)	.99 (.99, 1.0)
		90	.79 (.75, .83)	.64 (.63, .65)	.61 (.50, .72)	.99 (.99, 1.0)
		128	.77 (.76, .78)	.65 (.64, .65)	.69 (.66, .72)	.99 (.99, 1.0)
		360	.76 (.75, .78)	.62 (.61, .64)	.65 (.63, .66)	.99 (.99, .99)
		512	.76 (.76, .76)	.63 (.62, .63)	.62 (.59, .65)	.99 (.99, .99)
RANSynCoders	1×10^{-6}	45	.77 (.75, .79)	.62 (.58, .65)	.49 (.48, .5)	.70 (.64, .75)
		90	.76 (.73, .79)	.60 (.56, .64)	.50 (.49, .52)	.64 (.54, .74)
		180	.75 (.67, .83)	.55 (.46, .64)	.50 (.48, .53)	.70 (.67, .73)
		360	.76 (.73, .79)	.64 (.59, .69)	.49 (.47, .52)	.69 (.60, .78)
		720	.79 (.75, .82)	.65 (.62, .67)	.51 (.48, .54)	.63 (.56, .69)
	1×10^{-3}	45	.76 (.73, .78)	.64 (.60, .68)	.50 (.47, .53)	.66 (.56, .76)
		90	.77 (.75, .78)	.59 (.52, .65)	.53 (.50, .55)	.62 (.49, .75)
		180	.74 (.68, .80)	.61 (.59, .64)	.49 (.44, .53)	.67 (.62, .73)
		360	.78 (.75, .81)	.60 (.57, .63)	.51 (.49, .53)	.67 (.60, .74)
		720	.78 (.76, .80)	.62 (.60, .64)	.50 (.48, .52)	.66 (.59, .73)
GANF	1×10^{-6}	32	.73 (.68, .77)	.45 (.42, .47)	.46 (.37, .54)	.94 (.92, .97)
		64	.69 (.64, .75)	.53 (.51, .56)	.39 (.36, .43)	.94 (.92, .95)
		128	.66 (.50, .82)	.54 (.52, .57)	.49 (.47, .51)	.94 (.90, .99)
		256	.69 (.65, .73)	.51 (.47, .54)	.42 (.38, .46)	.80 (.75, .86)
		512	.67 (.66, .68)	.53 (.48, .58)	.53 (.47, .59)	.80 (.66, .94)
	1×10^{-3}	32	.54 (.40, .68)	.51 (.47, .55)	.58 (.48, .67)	.53 (.33, .74)
		64	.57 (.38, .76)	.46 (.43, .48)	.47 (.38, .56)	.69 (.51, .86)
		128	.71 (.63, .80)	.47 (.44, .50)	.49 (.42, .56)	.70 (.49, .91)
		256	.69 (.59, .78)	.53 (.51, .55)	.46 (.39, .52)	.69 (.47, .92)
		512	.68 (.56, .79)	.49 (.44, .53)	.50 (.43, .58)	.86 (.80, .93)

VI. CONCLUSION

Anomaly detection has grow in the past decade given prominence of machine learning models as well as the availability of high-quality acoustic data. In the industry, anomaly detection has been fundamental for predictive maintenance in industrial machinery, mitigating both economic and safety burdens. The framework MTSA proposed in this paper helps mitigate the replicability crisis in anomaly detection research, and makes it possible to fully understand a complex pipeline in terms of smaller and less-complex data processing building blocks. By implementing state-of-the-art approaches on MTSA, comparative analyses are facilitated, and data processing building blocks of complex pipelines can be reused into novel approaches.

Using MTSA, not only can comparative analyses between complex pipelines such as RANSynCoders and GANF be conducted, but novel approaches can also be developed by incorporating incremental improvements. Although RANSynCoders and GANF are not initially suited for acoustic data, once implemented on MTSA, it becomes straightforward to incorporate a feature extraction step based on MFCC at the beginning of their pipelines.

Beyond the parameters analyzed in this work, additional parameters can be explored and tested on various datasets, ensuring continuous refinement. For future works, we aim at improving the performance of the GANF and RANSynCoders models by modifying and varying their internal components.

By addressing the challenges faced in replicating machine learning models, especially in the context of anomaly detection using acoustic data, our work highlights the need for standardized practices to promote more replicability in anomaly detection research.

The development of MTSA as a unified framework, coupled with the ease of incorporating MFCC as a feature extraction method at the beginning of the RANSynCoders and GANF pipelines, has not only enabled these models to detect anomalies using acoustic data but also highlighted the merits of standardization and consistency in developing state-of-the-art models. This standardization ensures their reliability and replicability. In summary, MTSA aims to set higher standards in anomaly detection research, providing a solid foundation for developing trustworthy innovations applicable beyond the industrial sector.

ACKNOWLEDGEMENTS

Rodrigo Monteiro, Andrea Maria and Diego Pinheiro would like to thank the *STIC AmSud* (CAPES, Brazil) for financial support under grant number 001

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[2] R. de Paula Monteiro, M. C. Lozada, D. R. C. Mendieta, R. V. S. Loja, and C. J. A. B. Filho, "A hybrid prototype selection-based deep learning approach for anomaly detection in industrial machines," *Expert Systems with Applications*, vol. 204, p. 117528, 2022.

[3] Y. Chinniah, "Analysis and prevention of serious and fatal accidents related to moving parts of machinery," *Safety science*, vol. 75, pp. 163–173, 2015.

[4] L. F. M. Filho, R. d. P. Monteiro, D. Pinheiro, P. T. Endo, and A. M. N. C. Ribeiro, "Forecasting imminent failures in electrical industrial centrifuge using machine learning," in *2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pp. 1–6, 2023.

[5] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, "Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," 2019.

[6] H. Zhou, K. Yu, X. Zhang, G. Wu, and A. Yazidi, "Contrastive autoencoder for anomaly detection in multivariate time series," *Information Sciences*, vol. 610, pp. 266–280, 2022.

[7] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical Approach to Asynchronous Multivariate Time Series Anomaly Detection and Localization," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, (Virtual Event Singapore), pp. 2485–2494, ACM, Aug. 2021.

[8] E. Dai and J. Chen, "Graph-augmented normalizing flows for anomaly detection of multiple time series," in *2022 The International Conference on Learning Representations (ICLR)*, 2022.

[9] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, "Lstm-autoencoder based anomaly detection for indoor air quality time series data," 2022.

[10] I. Kobyzev, S. J. Prince, and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 3964–3979, 2021.

[11] D. PINHEIRO, "mtsa — pypi.org." https://pypi.org/project/mtsa/, 2024. [Accessed 29-05-2024].

[12] Z. K. Abdul and A. K. Al-Talabani, "Mel frequency cepstral coefficient and its applications: A review," vol. 10, pp. 122136–122158.