2021 Annual Reliability and Maintainability Symposium (RAMS) | 978-1-7281-8017-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/RAMS48097.2021.9605720

Deep Learning-based Anomaly Detection for Compressors Using Audio Data

Pooyan Mobtahej, Lamar University
Xulong Zhang, Lamar University
Maryam Hamidi, Lamar University
Jing Zhang, Lamar University

Key Words: Anomaly Detection, Midstream Infrastructure, Deep Learning

SUMMARY & CONCLUSIONS

By pressurizing natural gas in pipelines, the compression system interlocks upstream gas production and downstream consumer use. Considering the installation cost of $1 to $2 million US dollars for a compressor, the failure of the component can be costly. Therefore, the anomaly detection for the compression system is essential. In this paper, a deep learning-based anomaly detection method is proposed to identify the failure of midstream compressors using audio sensor data. Firstly, short-term Fourier transform (STFT), Mel-frequency cepstral coefficients (MFCC), and spectral centroid (SC) features are computed using the input audio signals. Secondly, deep learning-based feature extraction is applied to create high-level features. Finally, a principal component analysis step and a support vector machine are applied to classify normal and anomaly audio signals.

The proposed method was evaluated using two datasets with a total of 10196 audio signals collected from a compressor. The experimental results demonstrate that MFCC features are better than STFT features for anomaly detection and the combined deep MFCC features and SC features can achieve the best normal and anomaly signal classification performance, 100% for both datasets, using the proposed method.

1 INTRODUCTION

Compressor systems play an essential role in the connection of upstream gas production to downstream consumer use by pressurizing natural gas in pipelines. A compressor system typically includes a power unit, a cooling unit, and a compressor, which usually contains a crankcase, a valve body, and a turbo unit. Regular compression system inspection that monitors the working conditions of all components is very important because the failure of the compression system can be extremely expensive due to repair cost and lost production. However, traditional human-based inspection methods are very time-consuming and need experienced experts to listen to the compressor sounds. Moreover, the high decibel level of noise also prevents the accurate detection of abnormal sounds by the human ear. Therefore, a reliable system that can automatically analyze the auditory data, detect anomaly sounds, and predict component failure is urgently needed to solve this challenging problem.

Recently, sound sensors (e.g., microphone) have been widely used in automatic anomaly detection systems. Multiple sound sensors deployed at different components of a compressor system can be controlled by Internet of Things (IoT) techniques to collect and transfer the auditory data in time, and then advanced signal processing and machine learning techniques can be used to analyze the collected data and detect anomaly signals automatically.

A deep learning-based anomaly detection method for the compressor system using audio data is proposed in this paper. Short-term Fourier transform (STFT), Mel-frequency cepstral coefficients (MFCC), and spectral centroid (SC) features are first extracted, investigated and compared for normal and anomaly signals, and then a new deep learning-based method is presented to integrate MFCC high-level features and SC features for normal and anomaly audio signal classification. The experimental results demonstrate that the proposed method can classify all normal and anomaly signals correctly.

The rest of the paper is organized as follows: literature is reviewed in Section 2, the audio features and the proposed deep learning-based method are described in Section 3, Section 4 shows the experimental results and discusses the performance, and Section 5 draws the conclusions.

2 LITERATURE REVIEW

Much effort has been devoted to developing automatic anomaly detection systems using audio signals and many related publications can be found in the literature in recent years.

Prego *et al.* [1] used a three-stage algorithm for image processing to address anomaly detection problems. These stages consist of STFT, FEXT, and CLASS stages. The algorithm achieved decent anomaly detection rates on their dataset. Antonini *et al.* [2] used smart audio sensors for anomaly detection based on IoT architecture of received raw audio streams. Two algorithms adopted were Elliptic Envelope and Isolation Forest offered by Scikit-Learn. In AGILE -

978-1-7281-8017-5/21/$31.00 ©2021 IEEE

Adaptive Gateways for dIverse muLtiple Environments gateway instance, Elliptic Envelope performed better in terms of being faster and lighter. Erfani *et al.* [3] proposed DBN-1SVM hybrid model for resolving the challenge for anomaly detection in High-dimensional problem domains that is an unsupervised detection method, especially for large scale datasets. Combining these two methods as one hybrid model purpose some advantages such as resolving scalability issues with such complex and large datasets in detecting in training and feature detection. This model could also purpose a more accurate model, proper generalization and faster execution.

Deep learning technology has been growing and achieving impressive breakthroughs since 2012 in various research areas, such as artificial intelligence, computer vision, image understanding, and natural language process. Deep neural network models were also applied to detect and identify mechanical component failures.

Zhou *et al.* [4] presented a new method for anomaly detection. Their research aimed to improve a GAN network in terms of accuracy and better generalization. A pipeline structure with auto-encoder instead of the standard generator in GAN was used in the method. Koizumi *et al.* [5] developed an unsupervised method for unknown anomalous sound based on an autoencoder. A Neyman-Pearson lemma-based objective function was used to optimize the detection process. Rushe *et al.* [6] proposed a semi-supervised method for anomaly detection by assuming the anomalous patterns are not available for training. WaveNet architecture was applied to anomaly detection in raw audio. The experimental results showed that their method outperformed a based line autoencode model.

More publications about anomaly detection for IoT time-series and deep learning-based anomaly detection can be found in the two comprehensive survey papers [7] and [8].

3 METHODOLOGY

The goal of the proposed method is to classify the normal and anomaly audio signals collected from a compressor system. The STFT, MFCC, and SC features were compared and different methods were evaluated. The method using MFCC and SC features achieved the best performance and was selected as the proposed method in this paper.

Figure 1 shows the framework of the proposed method, which contains five steps. (1) MFCC features are computed using the input audio signal to form a two-dimensional feature matrix; (2) SC features are computed using the input audio signal; (3) A pre-trained deep learning neural network was used to extract high-level deep features from MFCC; (4) both deep MFCC features and SC features are fed to a principle component analysis (PCA) unit for feature extraction and dimension reduction; and (5) the extracted MFCC and SC features are combined to train a support vector machine (SVM) classifier for normal and anomaly audio signal classification.

3.1 *STFT Features*

A signal can be presented in both time domain and frequency domain. In the frequency domain, the signal can be decomposed into many constituent frequencies. STFT takes a short period of signal from a longer one using a window function, then applies Fourier transform to this short period of the signal [9]. This process is repeated from the start to the end of the original signal and the outputs of Fourier transform can be combined to generate a two-dimensional STFT image, which can show the frequency changes of the original signal.

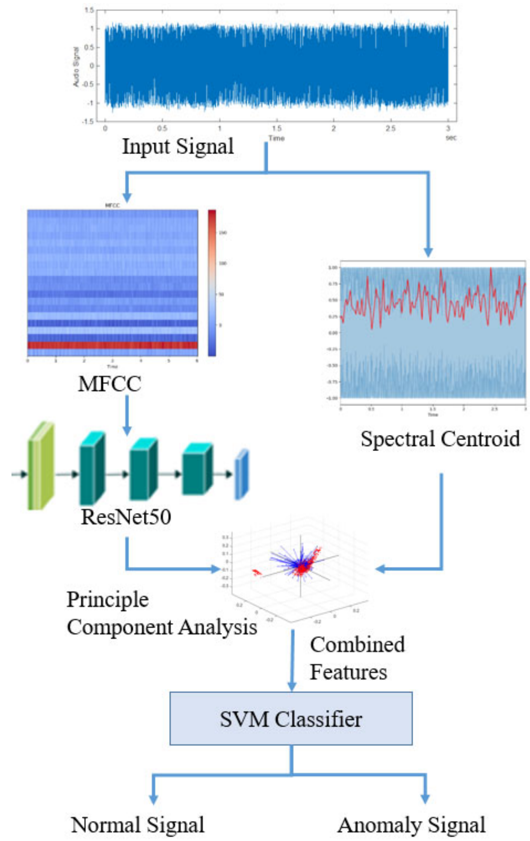

Fig. 1: The framework of the proposed method

The first row of Fig. 2 illustrates a normal and three anomaly audio signal waves and the second row shows their STFT images. The differences between the normal and the anomaly examples can be seen from the highlighted red boxes. Compared with the normal STFT, the STFTs of Anomaly-1 and Anomaly-2 both have narrow dark vertical patterns and there are clear horizontal textures in Anomaly-2's pattern. The SFTF of Anomaly-3 is very similar to the normal STFT except the dark rectangular regions in the lower half of the image.

3.2 *MFCC Features*

Mel-frequency analysis of an audio signal is based on human perception experiments. To concentrate on certain frequency components instead of the whole of the spectral envelop, a Mel filter bank with many filters are non-uniformly spaced on the frequency axis with more filters in the low frequency regions and less filters in the high frequency regions. The spectrum of the input audio signal after Fast Fourier Transform (FFT) is filtered by Mel-filters to create Mel-spectrum, and then the Cepstral analysis is performed to the logarithm of Mel-spectrum and the obtained Cepstral coefficients are referred to as Mel-Frequency Cepstral coefficients (MFCC) [10].

The third row of Fig. 2 shows four MFCC images corresponding to the input audio signals. It can be observed that the red regions at the bottom parts of the images are quite different. The normal example is smooth with no big value changes, while the anomaly examples have many low values embedded in between with big value changes.

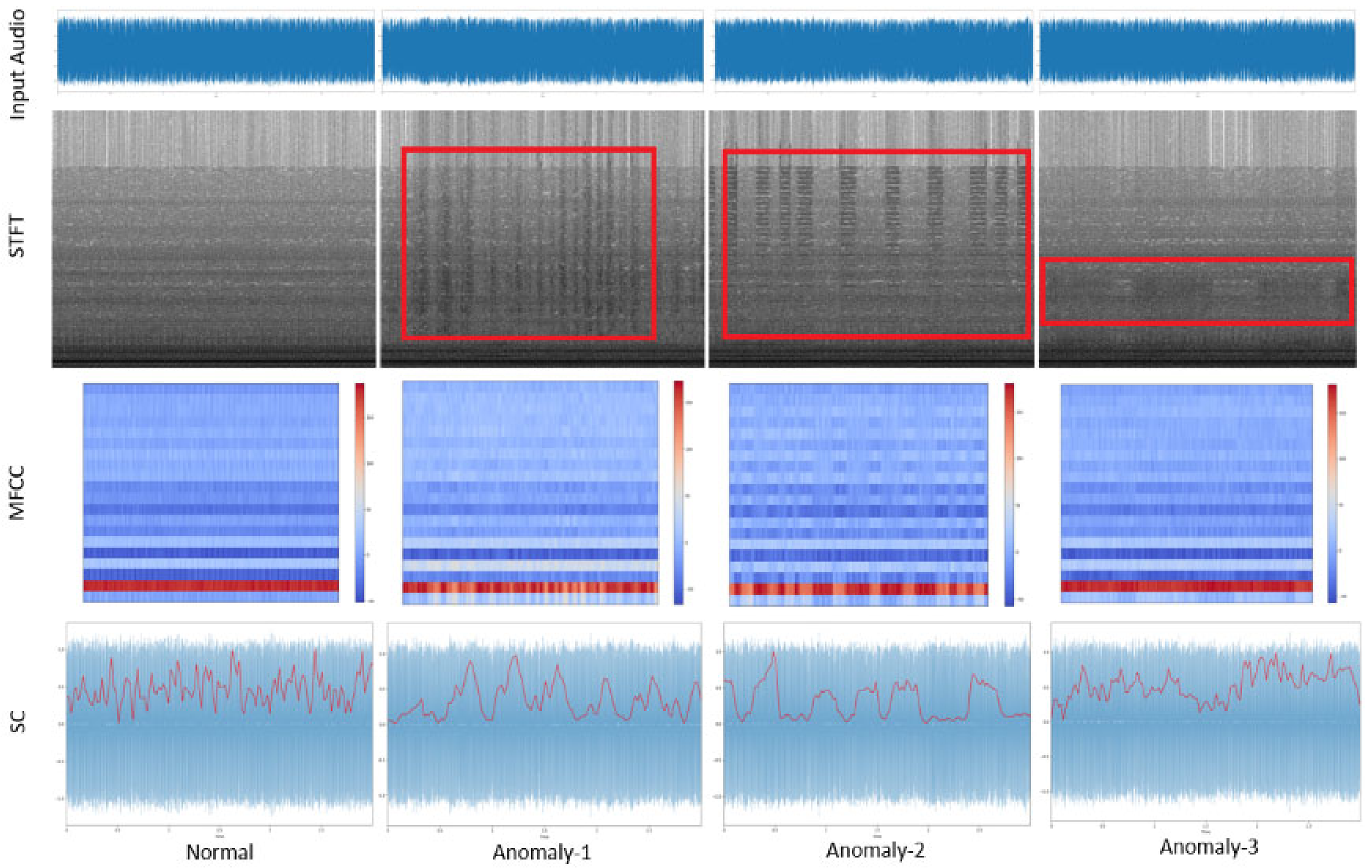

Fig. 2: A normal and three anomaly audio signals with their STFT, MFCC, and SC features

3.3 *Spectral Centroid Features*

Spectral Centroid (SC) [11] is another important feature for audio signal processing. The spectral centroid is associated with the brightness of a sound and indicates where the center of mass of the spectrum is. The individual centroid of a spectral frame can be calculated by using average frequency weighted amplitudes divided by the sum of amplitudes.

The computed SC signals of the four examples are shown in the last row of Fig. 2 in red color. Each SC signal can be considered as a 1×259 feature vector. Similar to STFT and MFCC features, the SC features show different properties for the normal and anomaly examples. The SC of the normal signal looks like a random signal with no clear periodic patterns, while the spectral centroids of the anomaly signals have periodic patterns that are consistent with the abnormal regions in their STFT images. For example, the SC features of Anomaly-1 and Anomaly-2 have short periods corresponding to the narrow vertical abnormal regions in their STFT images and the SC of Anomaly-3 has a long period corresponding to the wide rectangular abnormal regions in its STFT image.

3.4 *Deep Learning-based Feature Extraction*

Because STFT, MFCC, and SC features shown in Fig. 2 can capture the distinctive properties of the input audio signals, they can be used for normal and anomaly signal classification.

A deep convolutional neural network, ResNet50 model with a 50-layer architecture [12], was used to extract deep features from the two-dimensional STFT and MFCC images. ResNet50 is a residual network with the added shortcut connections that skip one or more layers and perform identity mapping, which allows deeper architectures to learn residuals left out by earlier layers rather than learn an unreferenced mapping. Furthermore, these identity mappings are parameter free thus reducing the number of flops. During backpropagation, the gradients pass through the identity mapping unaltered thus alleviating gradient vanishing.

The adopted ResNet50 was pre-trained using ImageNet, which is a well-known deep learning dataset containing over 14 million images from 1000 categories [13]. The network retraining or transfer learning was not applied due to the limited number of samples in the datasets used in this research. The high-level deep features were extracted from the average pooling layer after the last convolution layer, which was used as a feature extractor in several applications, such as weather classification, the detection of human rights violations, and parasite detection. The size of the extracted high-level deep features was 1×2048.

3.5 *PCA and SVM Classification*

PCA is widely used in data compression and redundant data removal. It applies an orthogonal linear transformation to find the big variances of the data and to make the data lie on the corresponding coordinates, which are called principal components. By selecting the first serval principal components, PCA can keep important information of the data while projects the data in a higher-dimensional space into a lower-dimensional space to reduce the data size.

SVM is one of the most used supervised classification models in the machine learning area. It classifies the data using optimal hyperplane that can maximize the margin between the two classes by finding support vectors. Different kernel functions can be used to map the data to a higher-dimensional space to make the problem linear separable.

In this research, the deep features extracted in Section 3.4 have 2048 dimensions, which is much higher than 259 dimensions of SC features. In order to integrate the extracted deep features and SC features and make them have equal contributions, both of them were input to the PCA and the first 30 principal components of each were used for classification. The output of PCA is a 60-dimensional feature set combined with the high-level deep STFT/MFCC features extracted by the ResNet50 model and the SC features. This feature set was then fed to a SVM classifier whose kernel function is a Gaussian function with a scale 7.7 to classify the audio signals into normal and anomaly classes.

4 *EXPERIMENTS*

4.1 *Dataset*

The following two datasets collected by Well Checked Systems International (http://www.wellchecked.com/) have been used to develop the proposed anomaly detection system.

(1) *Pettijohna* Dataset: This dataset has a total of 2343 raw audio signals collected on 15 Apr 2019. There are 2187 normal audio signals and 156 anomaly audio signals. Each signal is 3 seconds long and saved in the OGG format.

(2) *Pja* dataset: This dataset has a total of 7853 audio signals collected on 24 Apr 2019. There are 6768 normal audio signals and 1085 anomaly audio signals. Each signal is 3 seconds long and saved in the OGG format.

Both datasets are very unbalanced. There are many more normal samples than anomaly samples. Only 6.658% and 13.816% are anomaly audio signals, which increases the classification difficulty for anomaly samples.

4.2 *Evaluation Metrics and Experiment Settings*

To evaluate the anomaly detection performance of the proposed method, the following four evaluation metrics, *Accuracy*, *Precision*, *Recall*, and *F1* were used in the experiments.

$$Accuracy = \frac{TP + TN}{Total\ Number\ of\ Samples} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{Precison \times Recall}{Precision + Recall} \tag{4}$$

TP is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives, and *FN* is the number of false negatives. F1 is a combination of *Precision* and *Recall* and measures the overall performance. When both *Precision* and *Recall* are 1s, *F1* reaches the max value 1.

A 2-fold cross-validation was used to evaluate the performance using *Pettijohna* and *Pja* datasets. For each dataset, 50% of normal samples and 50% of anomaly samples were randomly selected and combined to create the training dataset and the remaining samples were used to create the test dataset. The performance was also evaluated by exchanging the training dataset and test dataset.

The audio signal feature extraction and ResNet50 model program was implemented using Python deep learning API Keras and PCA and SVM classification was implemented using MATLAB 2018b Classification Learner Toolbox. The computer used equipped with Intel Xeon CPU, 32 GB RAM, NVIDIA GeForce GTX 1080Ti GPU with 12 GB memory, and Debian GNU/Linux 10 (buster) OS

4.3 *Experimental Results*

A series of experiments were conducted to assess the extracted STFT, MFCC, SC features for anomaly detection using ResNet50 deep learning network comprehensively.

(1) STFT+ResNet50+SVM: The method that used STFT deep features extracted by ResNet50and SVM without the PCA feature extraction was considered as the baseline to compare with other methods and its experiment results are listed in Table 1. This method obtained high classification performance for both datasets.

Table 1: STFT+ResNet50+SVM

	Pettijohna dataset	*Pja* dataset
Accuracy	98.250%	95.976%
Precision	98.290%	95.629%
Recall	99.863%	99.897%
F1	99.070%	97.716%

(2) STFT+ResNet50+PCA+SVM: This method used STFT deep features extracted by ResNet50 and SVM with the PCA feature extraction.

Table 2: STFT+ResNet50+PCA+SVM

	Pettijohna dataset	*Pja* dataset
Accuracy	98.421%	96.396%
Precision	98.689%	96.065%
Recall	99.634%	99.911%
F1	99.407%	97.950%

The purpose of this experiment was to test the contribution of the PCA. The experimental results are listed in Table 2. It can be seen that this method achieved better performance compared with the baseline method because of the PCA applied.

(3) MFCC+ResNet50+PCA+SVM: The method used MFCC features extracted by ResNet50 and SVM with the PCA feature extraction. The experimental results are listed in Table 3. This method greatly improved performance compared to the first two methods. All four evaluation metrics for the two datasets are higher than 99%.

Table 3: MFCC+ResNet50+PCA+SVM

	Pettijohna dataset	*Pja* dataset
Accuracy	99.915%	99.211%
Precision	99.909%	99.107%
Recall	100%	99.985%
F1	99.954%	99.544%

(4) MFCC+ResNet50+SC+PCA+SVM: The method used MFCC deep features extracted by ResNet50 and SC features and SVM with the PCA feature extraction. The experimental results are listed in Table 4. All evaluation metrics reached 100% for the two datasets.

Table 4: MFCC+ResNet50+SC+PCA+SVM

	Pettijohna dataset	*Pja* dataset
Accuracy	100%	100%
Precision	100%	100%
Recall	100%	100%
F1	100%	100%

Please note that all values listed in Table 1 to Table 4 are the averages of the outputs of the 2-fold cross-validation described in Section 4.2.

4.4 *Discussions*

Based on the experimental results presented in Section 4.3, it can be seen that all tested methods yielded high classification performance and *Pettijohna* dataset is easier for classification than *Pja* dataset due to its small number of samples. Furthermore, there are other important research findings that are worth noting:

(1) Deep learning is a very powerful tool that can efficiently extract high-level deep features. One-dimensional audio signals can be converted to two-dimensional image signals to take advantage of the advanced deep learning technologies for audio signal analysis. The 2048-dimensional deep features extracted from two-dimensional STFT and MFCC images can capture the essential information of the input audio signals and achieve the high classification performances in this research.

(2) PCA can reduce the dimension of the deep features by extracting the key components and removing redundant information from the feature set. The normal and anomaly classification performance improved for both datasets as shown in Table 2. Therefore, the PCA can increase the anomaly detection ability for this research task.

(3) MFCC features are better than STFT features for the audio signal-based normal and anomaly classification task in this research. For example, by replacing STFT features with MFCC features, the *Accuracy*, *Precision*, *Recall*, and *F1* of *Pettijohna* dataset were improved by 1.494%, 1.220%, 0.366%, and 0.557%, respectively, and the *Accuracy*, *Precision*, *Recall*, and *F1* of *Pja* dataset were improved by 2.842%, 3.042%, 0.074%, and 1.594% respectively.

(4) The combination of multiple-resource features can improve the classification results in this research. Because MFCC and SC features extract different information from the input audio signals, the integration of MFCC and SC features can fuse useful information from different resources together to create a more comprehensive feature dataset, which can facilitate the classification step. The best performance was reached when using the combined MFCC+SC features. All evaluation metrics of the two datasets reached 100% as shown in Table 4.

5 CONCLUSIONS

Anomaly detection is critical in reducing the cost of midstream components maintenance. A deep learning-based anomaly detection method is presented in this paper to identify failures of the audio data collected from a compressor. MFCC, STFT, and SC features were computed using the input audio data, and then the high-level features extracted from MFCC/STFT by a deep learning network were fed to the PCA and SVM for classifying normal and anomalies classification. Two datasets with a total of 10196 audio signals were used to evaluate the proposed method. The experiments demonstrated that MFCC performed better than STFT for this research and the combined deep MFCC features and SC features achieved the best performance for normal and anomalies classification with 100% for four evaluation metrics used in the experiments.

The future work includes using other bigger audio signal datasets to test the performance of the proposed method, classifying the anomaly samples into sub-classes that correspond to different types of component failures, applying other deep learning networks (e.g., LSTM network) to analyze one-dimensional raw audio signals, and combining one-dimensional and two-dimensional audio signal features for anomaly detection.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Hayden Haines and Mr. Tim He from Well Checked Systems for providing the datasets used in this study. The authors also acknowledge the National Science Foundation for the high performance computing cluster at Lamar University supported under award No. 1726500. This research is supported by the Center for Midstream Management and Science at Lamar University.

REFERENCES

1. Prego, Thiago de M., Amaro A. de Lima, Sergio L. Netto, and Eduardo AB da Silva. "Audio anomaly detection on

rotating machinery using image signal processing." In *2016 IEEE 7th Latin American Symposium on Circuits & Systems (LASCAS)*, pp. 207-210. IEEE, 2016.
2. Antonini, Mattia, Massimo Vecchio, Fabio Antonelli, Pietro Ducange, and Charith Perera. "Smart audio sensors in the internet of things edge for anomaly detection." *IEEE Access* 6 (2018): 67594-67610.
3. Erfani, Sarah M., Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning." *Pattern Recognition* 58 (2016): 121-134.
4. Zhou, Jie, Yanxiang Chen, Pengcheng Zhao, and Shuanggen Fan. "Anomaly Detection with Pipeline Structure using Joint Distribution." In *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pp. 398-402. IEEE, 2020.
5. Koizumi, Yuma, Shoichiro Saito, Hisashi Uematsu, Yuta Kawachi, and Noboru Harada. "Unsupervised detection of anomalous sound based on deep learning and the Neyman–Pearson lemma." *IEEE/ACM Trans. on Audio, Speech, and Language Processing* 27, no. 1 (2018): 212-224.
6. Rushe, Ellen, and Brian Mac Namee. "Anomaly detection in raw audio using deep autoregressive networks." In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3597-3601. IEEE, 2019.
7. Chalapathy, Raghavendra, and Sanjay Chawla. "Deep learning for anomaly detection: A survey." *arXiv preprint arXiv:1901.03407* (2019).
8. Cook, Andrew, Göksel Mısırlı, and Zhong Fan. "Anomaly detection for IoT time-series data: A survey." *IEEE Internet of Things Journal* (2019).
9. Allen, Jonathan. "Short term spectral analysis, synthesis, and modification by discrete Fourier transform." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 25, no. 3 (1977): 235-238.
10. Zheng, Fang, Guoliang Zhang, and Zhanjiang Song. "Comparison of different implementations of MFCC." *Journal of Computer science and Technology* 16, no. 6 (2001): 582-589.
11. Peeters, Geoffroy. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project." *CUIDADO IST Project Report* 54, (2004): 1-25.
12. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.
13. Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248-255.

BIOGRAPHIES

Pooyan Mobtahej
Department of Industrial Engineering
Lamar University
4400 S M L King Jr Pkwy,
Beaumont, TX 77705 USA

e-mail: pmobtahej@lamar.edu

Pooyan is a Doctor of Industrial Engineering Student in Lamar university who had Engineering and Management background in Chalmers/Gothenburg University in Sweden. During his studies he has worked on several projects for reducing costs in industry which led to several industrial frameworks such as Innovation Maturity Model for reducing cost of development and production and improving efficiency. He is now working with data science and machine learning projects at LU.

Xulong Zhang
Department of Computer Science
Lamar University
4400 S M L King Jr Pkwy,
Beaumont, TX 77705 USA

e-mail: xzhang9@lamar.edu

Xulong Zhang is a graduate student in the Department of Computer Science at Lamar University. He earned his bachelor’s degree in communication engineering at Donghua University in Shanghai, China. He was a senior software engineer at Endeavor Streaming Shanghai office. His research interests include image processing, computer vision, machine learning, deep learning.

Maryam Hamidi, PhD
Department of Industrial Engineering
Lamar University
4400 S M L King Jr Pkwy,
Beaumont, TX 77705 USA

e-mail: mhamidi@lamar.edu

Maryam Hamidi earned her PhD in Systems and Industrial Engineering from the University of Arizona in 2016. She earned her MBA degree from Sharif University of Technology and her BS degree in Electrical Engineering from Amir-Kabir University of Technology, Iran. She joined the faculty at Lamar University in 2016 and is currently an Assistant Professor of the Industrial Engineering department and the Center for Advances in Port Management. Her areas of expertise are reliability, maintenance, port and waterway operations.

Jing Zhang, PhD
Department of Computer Science
Lamar University
4400 S M L King Jr Pkwy,
Beaumont, TX 77705 USA

e-mail: jing.zhang@lamar.edu

Jing Zhang is an assistant professor in the Department of Computer Science at Lamar University, Beaumont, TX. Dr. Zhang received his doctoral degree in Computer Science at the University of South Florida in 2012. He was a postdoc associate at Duke University 2012-2014. Dr. Zhang has published over 40 peer reviews journal and conference papers with more than 1300 worldwide citations. His research interests include image processing, computer vision, machine learning, deep learning, pattern recognition, and medical image analysis.