Anomalous Sound Detection with Machine Learning: A Systematic Review

Eduardo Carvalho Nunes

ALGORITMI Centre, Department of Information Systems, University of Minho, Guimarães, Portugal

Abstract. Anomalous sound detection (ASD) is the task of identifying whether the sound emitted from an object is normal or anomalous. In some cases, early detection of this anomaly can prevent several problems. This article presents a Systematic Review (SR) about studies related to Anomalous Sound Detection using Machine Learning (ML) techniques. This SR was conducted through a selection of 31 (accepted studies) studies published in journals and conferences between 2010 and 2020. The state of the art was addressed, collecting data sets, methods for extracting features in audio, ML models, and evaluation methods used for ASD. The results showed that the ToyADMOS, MIMII, and Mivia datasets, the Mel-frequency cepstral coefficients (MFCC) method for extracting features, the Autoencoder (AE) and Convolutional Neural Network (CNN) models of ML, the AUC and F1-score evaluation methods were most cited.

Keywords. Anomalous Sound Detection, Machine Learning, Systematic Review

1. Introduction

Anomaly Sound Detection (ASD) has received a lot of attention from the scientific machine learning community in recent years [1]. An anomaly in sound can indicate an error or defect, detecting the anomaly earlier can avoid a series of problems such as industrial equipment predictive maintenance and audio surveillance of roads [2].

Anomaly detection techniques can be categorized as supervised anomaly detection, semi-supervised anomaly detection, and unsupervised anomaly detection. Supervised anomaly detection requires the entire dataset to be labeled "normal" or "abnormal" and this technique is basically a type of binary classification task. Semi-supervised anomaly detection requires only data considered "normal" to be labeled, in this technique, the model will learn what "normal" data are like. Unsupervised anomaly detection involves unlabeled data. In this technique, the model will learn which data is "normal" and "abnormal" [3].

This paper presents a systematic review (SR) aiming to verify the state of the art in audio anomaly detection using machine learning techniques. Additionally, it was an-
analyzed which datasets, methods for extracting features in audio, ML models, and evaluation methods most used in the accepted primary studies. Thus, this survey can enable a general analysis of the scope of the work.

In addition to this introductory section, the paper is organized as follows. The Research Methodology section presents the concept of SR, the defined and used protocol, and the process of conducting the review. The “Results and discussion” section presents and discusses the results.

2. Research Methodology

Unlike traditional literature reviews, the SR is a rigorous and reliable research method that aims to select relevant research, collect and analyze data, and allow evaluation [5]. According to Kitchanhan’s suggestion, this paper was developed considering the 3 phases: planning, execution and analysis of results (Figure[1] [6]. In the planning phase, a protocol is defined specifying research questions, keywords, inclusion and exclusion criteria for primary studies and other topics of interest. In the execution phase, the bibliographic research is conducted following the defined protocol, it is in this phase that the inclusion and exclusion of primary studies is done. And finally, the results analysis phase, the extraction of the data is done and the results are compared.

![Figure 1. RS phases adapted [7].](image)

2.1. Planning

First, a detailed protocol was designed to describe the process and method to be applied in this SR (Table [1]). This protocol contains: objective, main question, keywords and synonyms, study language, sources search methods, study selection criteria, source list, quality form fields and data extraction form fields.

2.2. Selection

The searches were carried out between November and December 2020. Only recent studies (published since 2010) were considered for assessing the state of the art. The primary studies returned from the electronic database were identified through the search string:

("anomalous sound detection" OR "detecting anomalous audio" OR "detection of anomalous sound" OR "anomaly detection") AND ("machine learning" OR "supervised anomaly detection" OR "semi-supervised anomaly detection" OR "unsupervised anomaly detection") AND ("sound" OR "audio")
Table 1. Defined Protocol for this SR.

| **Objective** | This Systematic Literature Review Protocol (SLRP) presents the methodological structure for the implementation of the literature review stage on audio anomaly detection with machine learning techniques. |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| **Main Question** | What machine learning techniques for audio anomaly detection? |
| **Keywords and Synonyms** | Anomalous Sound Detection; Anomaly Detection; Detecting Anomalous Audio; Detection of Anomalous Sounds; Machine Learning; Self-Supervised Anomaly Detection; Semi-Supervised Anomaly Detection; Unsupervised Anomaly Detection; Audio; Sound. |
| **Study Language** | English. |
| **Sources Search Methods** | The sources should be available via the web, preferably in scientific databases in the area (computer science, computing, electronics). In addition to traditional databases, some can be included according to the results found. Primary studies in other media can also be selected, as long as they meet the requirements of the SR.

This process will be carried out by means of searches composed of keywords. Primary studies will be found from searches carried out on search portals for articles, theses, dissertations, and journals.

During the information retrieval procedure, the strings found preferably in Titles, Abstracts, and Keywords of each database will be considered.

After checking the relevance of the work, it will be selected for reading (full text). Primary studies will then be accepted or rejected. There will be (I) Inclusion and (E) Exclusion criteria for each primary study analyzed. |
| **Study Selection Criteria** | **Inclusion**: Anomaly detection in audio; uses machine learning technique; primary study is written in english.

**Exclusion**: Not detect anomaly in audio; not uses a machine learning technique; it is not written in english; full paper not found; not present the abstract; have a publication year outside the specified deadline (i.e., earlier than 2010). |
| **Sources** | In addition to the below sources, a search was made for papers in the research community on Detection and Classification of Acoustic Scenes and Events (DCASE).

ACM: [http://portal.acm.org](http://portal.acm.org)

IEEE: [https://ieeexplore.ieee.org/Xplore/home.jsp](https://ieeexplore.ieee.org/Xplore/home.jsp)

SCOPUS: [http://www.scopus.com](http://www.scopus.com)

DCASE2020: [http://dcase.community/](http://dcase.community/) |
| **Quality Form Fields** | Coherence and cohesion textual; Uses machine learning technique in an objective way; Machine learning techniques are cited; |
| **Extraction Form Fields** | Which Machine learning category? Anomaly detection category? Which dataset used? Which programming language is used? Which libraries or structure used? |
The selection process for primary studies is illustrated in Figure 2. In the first step, 3150 primary studies were identified. In the second step, the titles and abstracts were read and the inclusion and exclusion criteria were applied. In this step, 109 studies were accepted, 3002 studies were rejected and 38 studies were duplicated. In the third step, the introduction and conclusion were read and the inclusion and exclusion criteria were also used. In this step, 34 studies were accepted, 72 studies were rejected and 3 duplicated studies. In the fourth step, 25 studies fully were read and 25 studies were accepted. After completing the selection of studies, it was noted that DCASE was widely cited. With that, a manual search was made and 7 more studies were accepted.

2.3. Analysis of Results

This phase consists of a review and extraction of information. The Table 2 shows the number of primary studies collected in each indexed database. It is important to note that Scopus covers some results from the ACM and IEEE. For each primary study, a summary was written with the main study topics.
Table 2. Number of studies obtained in the indexed databases.

| Source | Nº of Studies | Accepted - Selection Phase | Accepted - Extraction Phase |
|--------|---------------|-----------------------------|----------------------------|
| ACM    | 376           | 10 (2.65%)                  | 2 (20%)                    |
| IEEE   | 1100          | 25 (2.27%)                  | 3 (12%)                    |
| SCOPUS | 1674          | 74 (4.42%)                  | 19 (25.67%)                |
| DCASE  | 49            | 7 (14.28%)                  | 7 (100%)                   |
| Total  | 3199          | 116 (3.62%)                 | 31 (26.72%)                |

In the selection phase, 116 studies were selected for the next phase. After completing the selection phase, 31 studies were selected for the extraction phase. It is important to note that the selection of studies related to DCASE was done manually, that the studies of the selection phase were all accepted for the extraction phase. The main topics of interest: Machine Learning Technique, Anomaly Type, Dataset, Audio Feature Extraction Method, Anomaly Detection Model, and Machine Learning Model Evaluation Method. The main results of SR are described in the results section.

2.4. StArt Tool

The StArt tool (State of the Art through Systematic Review) was used to support the SR process [22]. This tool was created by LaPES-Software Engineering Research Lab (Federal University of S˜ao Carlos) and was developed with the purpose of to automate the phases of SR. The tool offers full support for SR and is divided into: Planning, Execution, and Summary.

3. Results and Discussions

3.1. Journals and Proceedings

All primary studies were retrieved from scientific journals and conference proceedings. The Table 3 lists primary studies published in journals. In general, the journals affiliated to IEEE obtained more studies with 5 (50%) primary studies, two of which have the best impact factor. The Table 4 lists the primary studies published in Proceedings. As shown in the table, the DCASE 2020 contain more studies with 7 (33%) primary studies. It is important to highlight that in the event there was a competition (task 2) that is totally related to the detection of anomalies in audio. Another important observation is that the proceedings affiliated with IEEE, which had 8 primary studies (38%) accepted in this SR.

The studies were published by 133 different researchers. Table 8 (Appendix C) shows a summary of the researchers responsible for two or more studies. In this table we can see that the highlights are laboratories from Japan. Figure 4 (Appendix B) shows the total number (per country) of researchers with published studies. Highlights Japan with 38 (28%) researchers and Italy with 29 (22%) researchers.

Figure 3 shows the evolution of the research areas in relation to the number of published studies. According to this SR, the theme of this study had its first study published only in 2014. In the years 2017 and 2018, there were few published studies. However, in 2019 and 2020 many more studies began to emerge.
### Table 3. Journals that provided the included primary studies.

| Journal                                               | Impact Factor | N° of Studies |
|-------------------------------------------------------|---------------|---------------|
| IEEE Transactions on Intelligent Transportation Systems (ISSN 1558-0016) | 6.319         | 1             |
| IEEE Transactions on Information Forensics and Security (ISSN 1556-6021) | 6.013         | 1             |
| Engineering Applications of Artificial Intelligence (ISSN: 0952-1976) | 4.201         | 2             |
| IEEE Access (ISSN 2169-3536)                          | 3.745         | 1             |
| IEEE/ACM Transactions on Audio, Speech, and Language Processing (ISSN 2329-9304) | 3.398         | 2             |
| Electronics (ISSN 2079-9292)                          | 2.412         | 1             |
| Expert Systems (ISSN 1468-0394)                       | 1.546         | 1             |

### Table 4. Proceedings that provided the included primary studies.

| Proceeding                                                                 | N° of Studies |
|---------------------------------------------------------------------------|---------------|
| Proceedings of the Fifth Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE 2020) | 7             |
| IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) | 4             |
| IEEE Workshop on Evolving and Adaptive Intelligent Systems (EAIS)          | 1             |
| IEEE International Conference on Machine Learning and Applications (ICMLA) | 1             |
| IEEE Workshop on Machine Learning for Signal Processing                   | 1             |
| IEEE International Conference on Advanced Trends in Information Theory (ATIT) | 1             |
| International Conference on Industrial Engineering and Applications (ICIEA) | 1             |
| International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO) | 1             |
| European Signal Processing Conference (EUSIPCO)                           | 1             |
| International Conference on Intelligent Networking and Collaborative Systems (INCOS) | 1             |
| Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence | 1             |
| International Conference on Soft Computing and Machine Intelligence (ISCMI) | 1             |

#### 3.2. Primary studies accepted in the SR

Tables 5, 6, and 7 shows a synthesis of the 31 primary studies analyzed including: dataset, audio features, ML model and evaluation method for each one. A global analysis is presented in this session taking into account the most relevant topics.

About datasets, 9 studies created and used their own dataset, and 22 studies used
Table 5. Anomaly Detection in audio presented in the included studies of the SR (Automatic Search).

| Study Year | Dataset | Audio Features | ML Model | Evaluation Method |
|------------|---------|----------------|----------|-------------------|
| 2020       | Mivia Dataset [8] | STFT, MFCC, Mel-Scale | DenseNet-121, MobileNetV2, ResNet-50 | RR, MDR, ER, FPR, Accuracy |
| 2020       | ToyADMOS [11], MIMII [12] | Mel-Filterbank | SPIDERnet, AE, Naive MSE, PROTOnet | AUC, ROC, TPR, FPR, F-measure |
| 2020       | Mivia Dataset [8] | Audio Power, Audio harmonicity, Total loudness in Bark scale, Autocorrelation coefficient, ZCR, Log-attack time, Temporal centroid, Audio spectrum roll-off, Audio spectrum spread, Audio spectrum centroid, MFCC, Audio spectral flatness | one-class SVM, and DNN | Accuracy, F1-score, Precision |
| 2020       | Own Dataset | MFCC, and Mel filterbank energies | LSTM | Accuracy, and F1-score |
| 2020       | Own Dataset, Freesound [12] | MFCC, DWT, ZCR, SR, and GFCC | SVM, Random Forest, CNN, KNN Gradient Boosting, | Precision, Recall, F1-score, Accuracy, p-value |
| 2020       | Own Dataset | Mel-spectrogram | Conv-LSTM AE, and CAE | ROC-AUC, F1-score |
| 2020       | Own Dataset | Mel-spectrogram | CAE, and One-Class SVM | ROC-AUC, F1-score |
| 2020       | Mivia Dataset [8] | Gammatonogram images | AReN (CNN) | Accuracy, RR, MDR, ER, FPR |
| 2019       | Own Dataset | Mel-spectrogram | Deep AE | ROC-AUC curve |
| 2019       | Toy Car Running Dataset [13] | Time-series of acoustic | AE | AUC |
| 2019       | UrbanSound8K [14], TUT Dataset [15] | LPC, MFCC, and GFCC | Agglomerative Clustering, BIRCH | Precision, Recall, F1-score, TP, FP, FN |
| 2019       | TUT Dataset [17] | Raw audio | WaveNet, and CAE | ROC-AUC |
### Table 6. Continued. Anomaly Detection in audio presented in the included studies of the SR (Automatic Search).

| Study Year | Dataset | Audio Features | ML Model | Evaluation Method |
|------------|---------|----------------|----------|-------------------|
| 2019       | DCASE2018 Task 1 [12], DCASE2018 Task 2 [20] | FFT, and Log mel spectrogram | CNN      | F1-score, AUC, mAP, AP, ER |
| 2019       | TUT Dataset [15], NAB Data Corpus [21] | Raw data | One-Class SVM, and LSTM-AE | Accuracy |
| 2019       | Own Dataset | Log mel energy | Deep AE | AUC |
| 2019       | Own Dataset, Effects Library [16] | Log mel spectrum, MFCC, General Sound, i-vector | WaveNet, AE, BLSTM-AE, AR-LSTM | F1-score |
| 2019       | DCASE 2016 Dataset [17] | MFCC | AE, VAE, and VAEGAN | AUC, TPR, and pAUC |
| 2018       | General Sound Effects Library [16] | Log mel filterbank | WaveNet, AE, BLSTM-AE, AR-LSTM | F1-score |
| 2018       | A3FALL [18] | Log mel-energies, and DWT | Siamese NN, SVM, One-Class SVM | F1-score, Recall, Precision |
| 2018       | TUT Dataset [15] | MFCC | Elliptic Envelope, and Isolation Forest | F1-score, and ER |
| 2017       | Own Dataset | Mel-spectrogram | LSTM-AE | ROC |
| 2017       | Own Dataset, UrbanSound8K [14] | Mel-spectrogram, Gammatone filterbanks | KNN | ROC, and AUC |
| 2016       | Mivia Dataset [8] | MFCC, ZCR, Energy ratios in Bark sub-bands, Audio spectrum centroid, Audio spectrum roll-off, Audio spectrum spread | SVM | RR, MDR, ER, FPR, ROC, AUC, Sensitivity |
| 2014       | Own Dataset | ZCR, FFT, DWT, MFCC, | One-Class SVM | F1-score, SD, LPC, and LPCC |
public and private datasets. In total, 14 datasets and the main ones were identified, where the 3 most cited datasets were: ToyADMOS \[11\], MIMII \[12\], and Mivia \[8\]. The ToyADMOS dataset is a machine operating sound dataset that has approximately 540 hours of normal sound and approximately 12,000 hours of anomalous sound. ToyADMOS was designed to detect audio anomalies in research involving machine operation \[11\]. The MIMII dataset is a data set for investigation and inspection of defective industrial machines. It contains the sounds generated from four types of industrial machines (valves, pumps, fans and slide rails) \[12\]. Mivia dataset is an audio dataset composed of 6,000 events considered to be vigilance (glass break, shots and screams) \[8\].

In ML, features are the independent variables that serve as input to your ML system or model. The ML model uses features to learn and make predictions. About the audio features method, 34 methods were identified. The main methods of extracting features from the analyzed audio were: Mel-frequency cepstral coefficients (MFCCs), Log-Mel Energy, and Mel-spectrogram.

To answer the main question of this SR, 33 machine learning techniques were identified to detect anomalies in audio. However, two machine learning techniques stand out: the Autoencoder (AE) and the Convolutional Neural Network (CNN). In the most recent studies, the transfer learning method is being used. Transfer Learning is an ML method in which a model developed in one task is reused as a starting point in another task. The developed models identified in this study: DenseNet-121, MobileNetV2, and ResNet-50.

About the evaluation method, 23 (approximately 75%) studies used AUC-ROC and F1-score. AUC-ROC is a performance evaluation that involves classification problems with thresholds. AUC represents a degree of separability and ROC is a probability curve. The higher the AUC, the better the model for predicting a particular class. F1-score measures the accuracy of an ML model. It is widely used in classification, in our example, "normal" and "abnormal".
### Table 7. Anomaly Detection in audio presented in the included studies of the SR (Manual Search).

| Study Year | Dataset | Audio Features | ML Model | Evaluation Method |
|------------|---------|----------------|----------|-------------------|
| 2020       | ToyADMOS [11, 12] | Log-mel energies | ResNet   | AUC               |
| 2020       | ToyADMOS [11, 12] | Gammatone Spectrogram | AE       | ROC, AUC, pAUC    |
| 2020       | ToyADMOS [11, 12] | Spectrogram       | AE       | AUC, pAUC         |
| 2020       | ToyADMOS [11, 12] | Log-mel energies | CNN      | AUC, pAUC         |
| 2020       | ToyADMOS [11, 12] | Log-mel energies | CAE      | ROC, AUC, pAUC    |
| 2020       | ToyADMOS [11, 12] | Log-mel energies | ResNet, MobileNetV2, GroupMADE | AUC, pAUC |
| 2020       | ToyADMOS [11, 12] | Log-mel energies | CNN, PCA, RLDA, PLDA | AUC |

### 4. Conclusions

This paper presents a systematic review on anomaly detection in audio using machine learning techniques. This research had as main objective to obtain the state of the art, enabling an organization of ideas and summarization of information.

In total, 31 studies were selected to study machine learning techniques for anomaly sound detection. After the analysis, 33 machine learning techniques were identified, where AE and CNN were the most cited. We also analyzed the most-used datasets for anomaly detection as their respective methods for extracting features and the evaluation method for machine learning models.

It aims that this study, the result of a secondary study, may allow some direction in the works and research related to the theme. In particular, the author’s interest is related to the use of machine learning techniques for anomaly sound detection in-vehicle.
January 2021

References

[1] Kawaguchi, Y., & Endo, T. (2017, September). How can we detect anomalies from subsampled audio signals. In 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.

[2] Henze, D., Gorishi, K., Bruegge, B., & Simen, J. P. (2019, December). AudioForesight: A Process Model for Audio Predictive Maintenance in Industrial Environments. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 352-357). IEEE.

[3] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., & Vento, M. (2015). Audio surveillance of roads: A system for detecting anomalous sounds. IEEE transactions on intelligent transportation systems, 17(1), 279-288.

[4] Alla, S., & Adari, S. K. (2019). Beginning Anomaly Detection Using Python-Based Deep Learning. Apress.

[5] Almeida Biolchini, J. C., Mian, P. G., Natali, A. C. C., Conte, T. U., & Travassos, G. H. (2007). Scientific research ontology to support systematic review in software engineering. Advanced Engineering Informatics, 21(2), 133-151.

[6] Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. Journal of systems and software, 80(4), 571-583.

[7] dos Santos, A. C. C., Delamaro, M. E., & Nunes, F. L. (2013, May). The relationship between requirements engineering and virtual reality systems: A systematic literature review. In 2013 XV Symposium on Virtual and Augmented Reality (pp. 53-62). IEEE.

[8] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., & Vento, M. (2015). Reliable detection of audio events in highly noisy environments. Pattern Recognition Letters, 65, 22-28.

[9] Koizumi, Y., Saito, S., Uematsu, H., Harada, N., & Inomo, K. (2019, October). ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection. In 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (pp. 313-317). IEEE.

[10] Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. arXiv preprint arXiv:1909.09347.

[11] Sammarco, M., & Detyniecki, M. (2018). Crashzam: Sound-based Car Crash Detection. VEHITS.

[12] R. (2008, September 21). Pack: KITCHEN common sounds. Freesound.Org. https://freesound.org/people/Robinhood76/packs/3870/

[13] Yuma Koizumi, Shoichiro Saito, Masataka Yamaguchi, Shin Murata and Noboru Harada, ”Batch Uniformization for Minimizing Maximum Anomaly Score of DNN-based Anomaly Detection in Sounds,” in Proc of Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2019.

[14] Salamon, J., Jacoby, C., & Bello, J. P. (2014, November). A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 1041-1044).

[15] Aleksandri Diment, Annamaria Mesaros, Toni Heittola, & Tuomas Virtanen. (2018). TUT Rare sound events, Evaluation dataset [Data set]. Zenodo. http://doi.org/10.5281/zenodo.1160455

[16] “Series 6000 general sound effects library,” [http://www.sound-ideas.com/sound-effects/series-6000-sound-effects-library.html] [Accessed: 02- Jan- 2021]

[17] DCASE2016 Challenge - DCASE. (2016). http://dcase.community/challenge2016/index

[18] Principi, E., Droghini, D., Squartini, S., Olivetti, P., Piazza, F. (2016). Acoustic cues from the floor: a new approach for fall classification. Expert Systems with Applications, 60, 51-61.

[19] A. Mesaros, T. Heittola, and T. Virtanen, “A multi-device dataset for urban acoustic scene classification,” in Proc. Detect. Classif. Acoust. Sce. Events 2018 Workshop, 2018, pp. 9–13

[20] E. Fonseca et al., “General-purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline,” 2018 arXiv:1807.09902

[21] Various artificial signal datasets, https://github.com/numenta/NAB/tree/master/data

[22] Fabbri, S., Hernandez, E., Di Thommazo, A., Belgamo, A., Zamboni, A., & Silva, C. (2012, June). Using information visualization and text mining to facilitate the conduction of systematic literature reviews. In International Conference on Enterprise Information Systems (pp. 243-256). Springer, Berlin, Heidelberg.

[23] Papadimitriou, I., Vafiadis, A., Lallas, A., Votis, K., & Tzovaras, D. (2020). Audio-Based Event Detection at Different SNR Settings Using Two-Dimensional Spectrogram Magnitude Representations. Electronics, 9(10), 1593.
[24] Koizumi, Y., Yasuda, M., Murata, S., Saito, S., Uematsu, H., & Harada, N. (2020, May). SPIDERnet: Attention Network For One-Shot Anomaly Detection In Sounds. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 281-285). IEEE.

[25] Rovetta, S., Maasri, Z., & Masulli, F. (2020, May). Detection of hazardous road events from audio streams: An ensemble outlier detection approach. In 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (E AIS) (pp. 1-6). IEEE.

[26] Becker, P., Roth, C., Roennau, A., & Dillmann, R. (2020, April). Acoustic Anomaly Detection in Additive Manufacturing with Long Short-Term Memory Neural Networks. In 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA) (pp. 921-926). IEEE.

[27] Vafeiadis, A., Votis, K., Giakoumis, D., Tzovaras, D., Chen, L., & Hamzaoui, R. (2020). Audio content analysis for unobtrusive event detection in smart homes. Engineering Applications of Artificial Intelligence, 89, 103226.

[28] Bayram, B., Duman, T. B., & Ince, G. (2020). Real time detection of acoustic anomalies in industrial processes using sequential autoencoders. Expert Systems, e12564.

[29] Koizumi, Y., Saito, S., Yamaguchi, M., Murata, S., & Harada, N. (2019, October). Batch uniformization for minimizing maximum anomaly score of dnn-based anomaly detection in sounds. In 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (pp. 6-10). IEEE.

[30] Komatsu, T., Hayashiy, T., Kondo, R., Todaz, T., & Takeda, K. (2019, May). Scene-dependent Anomalous sound event detection based on conditional wavenet and i-vector. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 870-874). IEEE.

[31] Henze, D., Gorishi, K., Bruegge, B., & Simen, J. P. (2019, December). AudioForesight: A Process Model for Audio Predictive Maintenance in Industrial Environments. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 352-357). IEEE.

[32] Kawaguchi, Y., Tanabe, R., Endo, T., Ichige, K., & Hamada, K. (2019, May). Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 865-869). IEEE.

[33] Komatsu, T., Hayashiy, T., Kondo, R., Todaz, T., & Takeday, K. (2019, May). Anomalous Sound Event Detection Based On Conditional Wavenet and I-vector. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 870-874). IEEE.

[34] Rushe, E., & Mac Namee, B. (2019, May). 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.

[35] Kawaguchi, Y., Tanabe, R., Endo, T., Ichige, K., & Hamada, K. (2019, May). Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 865-869). IEEE.

[36] Komatsu, T., Hayashiy, T., Kondo, R., Todaz, T., & Takeday, K. (2019, May). Scene-dependent Anomalous Sound Event Detection Based on Conditional Wavenet and I-vector. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 870-874). IEEE.

[37] Koizumi, Y., Saito, S., Uematsu, H., Kawachi, Y., & Harada, N. (2018). Unsupervised detection of anomalous sound based on deep learning and the Neyman-Pearson lemma. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(1), 212-224.

[38] Hayashi, T., Komatsu, T., Kondo, R., Toda, T., & Takeda, K. (2018, September). Anomalous sound event detection based on wavenet. In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 2494-2498). IEEE.

[39] Kawaguchi, Y., & Endo, T. (2017, September). How can we detect anomalies from subsampled audio signals?. In 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.

[40] Marchegiani, L., & Posner, I. (2017, May). Leveraging the urban soundscape: Auditory perception for smart vehicles. In 2017 IEEE International Conference on Robotics and Automation (ICRA) (pp. 6547-6554). IEEE.

[41] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., & Vento, M. (2015). Audio surveillance of roads: A system for detecting anomalous sounds. IEEE transactions on intelligent transportation systems, 17(1), 279-288.

[42] Aurino, F., Folla, M., Gargiulo, F., Moscato, V., Picariello, A., & Sansone, C. (2014, September). One-class SVM based approach for detecting anomalous audio events. In 2014 International Conference on Intelligent Networking and Collaborative Systems (pp. 145-151). IEEE.
[43] Droghini, D., Vesperini, F., Principi, E., Squarini, S., & Piazza, F. (2018, August). Few-shot siamese neural networks employing audio features for human-fall detection. In Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence (pp. 63-69).

[44] Kong, Q., Xu, Y., Sobieraj, I., Wang, W., & Plumbley, M. D. (2019). Sound event detection and time–frequency segmentation from weakly labelled data. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(4), 777-787.

[45] Provotar, O. I., Linder, Y. M., & Veres, M. M. (2019, December). Unsupervised Anomaly Detection in Time Series Using LSTM-Based Autoencoders. In 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT) (pp. 513-517). IEEE.

[46] Antonini, M., Vecchio, M., Antonelli, F., Ducange, P., & Perera, C. (2018). Smart audio sensors in the internet of things edge for anomaly detection. IEEE Access, 6, 67594-67610.

[47] Primus, P., Haenschmid, V., Praher, P., & Widmer, G. (2020). Anomalous Sound Detection as a Simple Binary Classification Problem with Careful Selection of Proxy Outlier Examples. arXiv preprint arXiv:2011.02949.

[48] Perez-Castanos, S., Naranjo-Alcazar, J., Zuccarello, P., & Cobos, M. (2020). Anomalous Sound Detection using unsupervised and semi-supervised autoencoders and gammatone audio representation. arXiv preprint arXiv:2006.15321.

[49] Park, J., & Yoo, S. DCASE 2020 TASK2: ANOMALOUS SOUND DETECTION USING RELEVANT SPECTRAL FEATURE AND FOCUSING TECHNIQUES IN THE UNSUPERVISED LEARNING SCENARIO.

[50] Inoue, T., Vinayavekhin, P., Morikuni, S., Wang, S., Trong, T. H., Wood, D., ... & Tachibana, R. (2020). Detection of Anomalous Sounds for Machine Condition Monitoring using Classification Confidence (Vol. 2). Tech. report in DCASE2020 Challenge Task.

[51] Kapka, S. (2020). ID-Conditioned Auto-Encoder for Unsupervised Anomaly Detection. arXiv preprint arXiv:2007.05314.

[52] Giri, R., Tenneti, S. V., Cheng, F., Helwani, K., Isik, U., & Krishnaswamy, A. SELF-SUPERVISED CLASSIFICATION FOR DETECTING ANOMALOUS SOUNDS.

[53] Wilkinghoff, K. USING LOOK, LISTEN, AND LEARN EMBEDDINGS FOR DETECTING ANOMALOUS SOUNDS IN MACHINE CONDITION MONITORING.
A. Image - Number of authors per year

![Bar chart showing the total number of authors per country](chart)

**Figure 4.** Number of primary authors per year.

B. Table 8 - Researchers of studies
Table 8. Researchers (author and co-author) and their publications contained in this SR.

| Author          | Pub | Affiliation                                                      | Country        |
|-----------------|-----|----------------------------------------------------------------|----------------|
| Koizumi, Y.     | 4   | NTT Media Intelligence Laboratories                             | Japan          |
| Saito, S.       | 3   | NTT Media Intelligence Laboratories                             | Japan          |
| Uematsu, H.     | 3   | NTT Media Intelligence Laboratories                             | Japan          |
| Harada, N.      | 3   | NTT Media Intelligence Laboratories                             | Japan          |
| Vafeiadis, A.   | 2   | Center for Research and Technology Hellas-Information Technologies Institute | Greece         |
| Votis, K.       | 2   | Center for Research and Technology Hellas-Information Technologies Institute | Greece         |
| Tzovaras, D.    | 2   | Center for Research and Technology Hellas-Information Technologies Institute | Greece         |
| Saggese, A.     | 2   | Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno | Italy          |
| Vento, M.       | 2   | Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno | Italy          |
| Vecchio, M.     | 2   | OpenIoT research unit, FBK CREATE-NET                            | Italy          |
| Antonini, M.    | 2   | OpenIoT research unit, FBK CREATE-NET                            | Italy          |
| Antonelli, F.   | 2   | OpenIoT research unit, FBK CREATE-NET                            | Italy          |
| Murata, S.      | 2   | NTT Media Intelligence Laboratories                             | Japan          |
| Kawaguchi, Y.   | 2   | Research and Development Group, Hitachi                         | Japan          |
| Endo, T.        | 2   | Research and Development Group, Hitachi                         | Japan          |
| Komatsu, T.     | 2   | Data Science Research Laboratories, NEC Corporation              | Japan          |
| Hayashi, T.     | 2   | Department of Information Science, Nagoya University             | Japan          |
| Kondo, R.       | 2   | Data Science Research Laboratories, NEC Corporation              | Japan          |
| Toda, T.        | 2   | Information Technology Center, Nagoya University                 | Japan          |
| Takeda, K.      | 2   | Department of Information Science, Nagoya University             | Japan          |
| Petkov, N.      | 2   | Faculty of Science and Engineering, University of Groningen      | Netherlands    |
| Bayram, B.      | 2   | Faculty of Computer and Informatics Engineering, Istanbul Technical University | Turkey         |
| Duman, T.B.     | 2   | Faculty of Computer and Informatics Engineering, Istanbul Technical University | Turkey         |
| Ince, G.        | 2   | Faculty of Computer and Informatics Engineering, Istanbul Technical University | Turkey         |
Table 9. Continuation of Table 4 (part 1)

| Author           | Pub | Affiliation                                                                                     | Country          |
|------------------|-----|------------------------------------------------------------------------------------------------|------------------|
| Plumbley, M. D.  | 1   | Centre for Vision, Speech and Signal Processing University of Surrey                              | England          |
| Becker, P.       | 1   | FZI Research Center for Information Technology                                                   | Germany          |
| Roth, C.         | 1   | FZI Research Center for Information Technology                                                   | Germany          |
| Roennau, A.      | 1   | FZI Research Center for Information Technology                                                   | Germany          |
| Dillmann, R.     | 1   | FZI Research Center for Information Technology                                                   | Germany          |
| Henze, D.        | 1   | Technische Universität München                                                                  | Germany          |
| Gorishti, K.     | 1   | Technische Universität München                                                                  | Germany          |
| Bruegge, B.      | 1   | Technische Universität München                                                                  | Germany          |
| Simen, J.-P.     | 1   | Carl Zeiss AG                                                                                   | Germany          |
| Stemmer, G.      | 1   | Intel Corp, Intel Labs                                                                          | Germany          |
| Rushe, E.        | 1   | Insight Centre for Data Analytics, University College Dublin                                    | Ireland          |
| Namee, B.M.      | 1   | Insight Centre for Data Analytics, University College Dublin                                    | Ireland          |
| Rovetta, S.      | 1   | DIBRIS, Universita degli studi di Genova                                                          | Italy            |
| Mnasri, Z.       | 1   | DIBRIS, Universita degli studi di Genova                                                          | Italy            |
| Masulli, F.      | 1   | DIBRIS, Universita degli studi di Genova                                                          | Italy            |
| Greco, A.        | 1   | Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno | Italy            |
| Janjua, Z.H.     | 1   | OpenIoT research unit, FBK CREATE-NET                                                             | Italy            |
| Foggia, P.       | 1   | Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno | Italy            |
| Strisciuglio, N. | 1   | Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno | Italy            |
| Aurino, F.       | 1   | Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II | Italy            |
| Folla, M.        | 1   | Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II | Italy            |
| Gargiulo, F.     | 1   | Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II | Italy            |
| Moscato, V.      | 1   | Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II | Italy            |
| Picariello, A.   | 1   | Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II | Italy            |
| Sansone, C.      | 1   | Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II | Italy            |
| Droghini, D.     | 1   | Università Politecnica delle Marche                                                              | Italy            |
| Vesperini, F.    | 1   | Università Politecnica delle Marche                                                              | Italy            |
| Principi, E.     | 1   | Università Politecnica delle Marche                                                              | Italy            |
| Squartini, S.    | 1   | Università Politecnica delle Marche                                                              | Italy            |
| Piazza, F.       | 1   | Università Politecnica delle Marche                                                              | Italy            |
| Ducange, P.      | 1   | SMARTEST Research Centre, eCampus University                                                      | Italy            |
| Yasuda, M.       | 1   | NTT Media Intelligence Laboratories                                                              | Japan            |
| Yamaguchi, M.    | 1   | NTT Media Intelligence Laboratories                                                              | Japan            |
| Tanabe, R.       | 1   | Research and Development Group, Hitachi                                                           | Japan            |
Table 10. Continuation of Table 4 (part 2)

| Author       | Pub | Affiliation                                      | Country          |
|--------------|-----|--------------------------------------------------|------------------|
| Ichige, K.   | 1   | Research and Development Group, Hitachi          | Japan            |
| Hamada, K.   | 1   | Research and Development Group, Hitachi          | Japan            |
| Inoue, T.    | 1   | IBM Research                                    | Japan            |
| Vinayavekhin, P. | 1 | IBM Research                                   | Japan            |
| Morikuni, S. | 1   | IBM Research                                    | Japan            |
| Tachibana, R. | 1 | IBM Research                                   | Japan            |
| Lopez-Meyer, P. | 1 | Intel Corp, Intel Labs                         | Mexico           |
| Kapka, S.    | 1   | Samsung R and D Institute Poland                | Poland           |
| Park, J.     | 1   | Advanced Robot Research Laboratory, LG Electronics | South Korea    |
| Yoo, S.      | 1   | Advanced Robot Research Laboratory, LG Electronics | South Korea    |
| Perez-Castanos, | 1 | Visualfy, Benisano                            | Spain            |
| Naranjo-Alcazar, J. | 1 | Visualfy, Benisano                          | Spain            |
| Zuccarello, P. | 1 | Visualfy, Benisano                            | Spain            |
| Cobos, M.    | 1   | Universitat de Valencia                         | Spain            |
| Provotar, O. I. | 1 | Faculty of Computer Science and Cybernetics Taras Shevchenko National University of Kyiv | Ukraine |
| Linder, Y. M. | 1 | Faculty of Computer Science and Cybernetics Taras Shevchenko National University of Kyiv | Ukraine |
| Veres, M. M. | 1   | Faculty of Computer Science and Cybernetics Taras Shevchenko National University of Kyiv | Ukraine |
| Giri, R.     | 1   | Amazon Web Services                             | USA              |
| Tenneti, S. V. | 1 | Amazon Web Services                           | USA              |
| Cheng, F.    | 1   | Amazon Web Services                             | USA              |
| Helwani, K.  | 1   | Amazon Web Services                             | USA              |
| Isik, U.     | 1   | Amazon Web Services                             | USA              |
| Krishnaswamy, A. | 1 | Amazon Web Services                       | USA              |
| Wang, S.     | 1   | IBM Research                                   | USA              |
| Trong, T. H. | 1   | IBM Research                                   | USA              |
| Wood, D.     | 1   | IBM Research                                   | USA              |
| Lopez, J. A. | 1   | Intel Corp, Intel Labs                         | USA              |
| Lu, H.       | 1   | Intel Corp, Intel Labs                         | USA              |
| Nachman, L.  | 1   | Intel Corp, Intel Labs                         | USA              |
| Huang, J.    | 1   | Intel Corp, Intel Labs                         | USA              |
| Perera, C.   | 1   | School of Computer Science and Informatics,     | Wales            |
|              |     | Cardiff University                              |                  |
| Primus, P.   | 1   | Institute of Computational Perception          | Austria          |
| Haunschmid, V. | 1 | Institute of Computational Perception        | Austria          |
| Praher, P.   | 1   | Software Competence Center Hagenberg GmbH      | Austria          |
| Widmer, G.   | 1   | LIT Artificial Intelligence Lab, Johannes Kepler University | Austria |
| Chen, L.     | 1   | Faculty of Computing, Engineering and Media, De Montfort University | England |
| Hamzaoui, R. | 1   | Faculty of Computing, Engineering and Media, De Montfort University | England |
| Marchegiani, L. | 1 | Oxford Robotics Institute, University of Oxford | England |
| Papadimitriou, I. | 1 | Center for Research and Technology Hellas-Information Technologies Institute | Greece |
| Lalas, A.    | 1   | Center for Research and Technology Hellas-Information Technologies Institute | Greece |
Table 11. Continuation of Table 4 (part 3)

| Author   | Pub | Affiliation                                      | Country  |
|----------|-----|--------------------------------------------------|----------|
| Posner, I. | 1   | Oxford Robotics Institute, University of Oxford  | England  |
| Kong, Q. | 1   | Centre for Vision, Speech and Signal Processing University of Surrey | England  |
| Xu, Y.   | 1   | Centre for Vision, Speech and Signal Processing University of Surrey | England  |
| Sobieraj, I. | 1 | Centre for Vision, Speech and Signal Processing University of Surrey | England  |
| Wang, W. | 1   | Centre for Vision, Speech and Signal Processing University of Surrey | England  |