SN Computer Science (2024) 5:542
https://doi.org/10.1007/s42979-024-02844-y

ORIGINAL RESEARCH

Self-Adaptive Incremental PCA-Based DBSCAN of Acoustic Features for Anomalous Sound Detection

Xiao Tan[1] · Siu Ming Yiu[1]

Received: 16 December 2023 / Accepted: 27 March 2024

Abstract

In modern industry, maintaining continuous machine operations is important for improving production efficiency and reducing costs. Therefore, the smart technology of acoustic monitoring to detect anomalous machine conditions earlier before breakdowns works as part of predictive maintenance and is applied not only in industry fault detection but also in safety monitoring and surveillance systems. This paper proposes a self-adaptive unsupervised machine learning algorithm with dimension-reduction technology to detect anomalous sounds after extracting acoustic machine features. Technically, the automatic EPS calculation algorithm-based genetic algorithm optimizes the automatic clustering algorithm's configuration for incremental principal component analysis and density-based spatial clustering algorithms with noise. IPCA is enhanced by the sequential Karhunen–Loeve (SKL) algorithm, and the condensation algorithm works as the second layer of the algorithm to reduce the number of effective components. This architecture could select an optimized set of parameters based on different test environments and keeps performance quality with fewer computational requirements. In the experiments, 228 sets of normal sounds and 100 sets of anomaly sounds are used. The sound files are collected from the same machine type (stepper motors) at a real plant site. We compare the proposed algorithm with K-means++, one-class SVM, agglomerative clustering, DCGAN and DCNN-Autoencoder, and this new algorithm performs best, with an AUC of 0.84 and the shortest execution time. The algorithm is generic and can be applied to detect anomalies in machines to provide early warning to people to avoid serious accidents or disasters.

Keywords Internet of things (IoT) · Anomalous sound detection · Unsupervised algorithm · Acoustic features · Guided genetic algorithm · Dimension reduction

Introduction

The industrial Internet of things (IoT), as an industry 4.0 implementation technology [1], is used in manufacturing to control and monitor operations and processes by smart sensors that detect the anomalous behaviors of machines, remotely control the input and output of each step in the process, and integrate physical production into interconnected networks.

Anomalous sound detection (ASD) is a smart data-driven technology at the edge of the IoT. Scientific methodology is used to identify the anomalous sound emitted from operational machines, and the detected warnings are sent to operators to mitigate the risk of breakdowns. For example, the modern textile industry uses a wide range of machines, especially massive heavy-duty industrial machines, e.g., woolen mill machines, thread winding machines, bleaching/dyeing machines, and scutching machines. The costs of detecting and fixing defects in those running machines in time are high, not only due to the expensive repair charges but also downtime. ASD, as an option for predictive maintenance technology, can detect fault conditions and automatically report them to operators in real time.

In addition to reducing the maintenance cost of audio analysis, anomaly detection technology can also be applied to image, video and text analysis in traffic control, cybersecurity and forensics. For example, in the automotive industry,

This article is part of the topical collection "Advances in Applied Informatics" guest edited by Hector Florez, Olmer Garcia and Florencia Pollo-Cattaneo.

✉ Xiao Tan
xtan@cs.hku.hk

Siu Ming Yiu
smyiu@hku.hk

[1] Department of Computer Science, University of Hong Kong, Pokfulam Road, Hong Kong, China

Published online: 13 May 2024

machine learning and artificial intelligence technology are adopted to recognize traffic lights using onboard sensors in vehicles to improve the safety of driving [2]. In utilities, AI-based smart sensors and anomaly detection methods are also widely used in traffic flow studies to improve the mobility of cities or crossing regions [3]. In cybersecurity, to detect anomalies, machine learning methods are applied to reduce the vulnerability of sensors, e.g., IoT-based smart grids (SGs) [4]. In forensics, anomaly detection with autonomous artificial intelligence is used to detect frauds or cybercrimes. AI-based anomaly detection technology is used to detect malicious and illicit events in the text analysis of posts in online social networks in dark web environments [5]. In addition, by analyzing the security logs of attacked servers, anomaly detection technology can help engineers trace threat-intention cyber behaviors and predict evidential locations [5]. Smart anomaly detection technology is a prominent approach in system automation and risk control in both industry and society.

However, ASD has become increasingly challenging in recent decades, despite the wide recognition of its importance in industry 4.0. The major challenges in practice include the following:

Imbalanced training dataset In practical applications, anomaly events are much rarer than long time series of normal data [6]. Such an imbalance between exhaustive continuous normal data and anomalous data in the training process significantly compromises the performance of popular ASD machine learning algorithms.

Stability of high performance Maintaining a stable and highly accurate detection and prediction performance is another issue in real practice. Most deep learning algorithms, e.g., convolutional generative adversarial networks (GANs), can achieve high accuracy after sufficient training. However, the stability of the overall predictive performance is still a concern [7].

Hardcoded architecture Differences in background environments when collecting sounds and types of sounds require different parameter settings in the algorithm. Manually selecting the parameters to reset the algorithm to adapt to the environment and specific types of sound impacts the efficiency and accuracy of ASD.

Noise On most occasions, the real environments in which sound data are collected are composed of multiple types of sound. Environmental noise is a traditional issue in audio studies [8].

High computation capability and computing cost requirements Because of the high volume of the training dataset and the imbalance between normal and anomalous data, the algorithms applied in ASD, e.g., deep convolutional neural networks and generative adversary networks, require one or more graphics processing units to process and generate good predictive results.

To resolve these issues in practice, the proposed algorithm integrates the dimension-reduction technology of incremental PCA with unsupervised DBSCAN. This algorithm is optimized with the automatic EPS calculation (AEC)-guided genetic algorithm to set the localized parameters for different test datasets [9]. The details of the algorithm are introduced in "Enhanced Incremental Principal Component Analysis-Based Density-Based Spatial Clustering of Applications with Noise".

We extend our gratitude to Mr. Huang CS, who provided the audio files for the study, and the Department of Computer Science at the University of Hong Kong, who sponsored the study.

Enhanced Incremental Principal Component Analysis-Based Density-Based Spatial Clustering of Applications with Noise

Extraction of Acoustic Features

Acoustic features are used to represent and recognize a typical computationally sound event or scenario to differentiate it from others. The input, as of the discrete time-series audio data of machine sounds collected from the plant site, is analog–digital converted, framed and partly labeled in the preprocessing stage and then is calculated and output as acoustic features by the preset rules. These digitalized representations, or acoustic characteristics, are capable of identifying the physical properties of the input audio data, for example, the signal energy, the toneless, the temporal shape and the spectral shape.

In recent decades, many different types of audio signal features have been proposed for sound recognition or description. Generally, the audio features can be categorized as either time domain or frequency domain. In the time domain, based on the different computational scopes, we can distinguish between the time extension validity of the global descriptors that are computed for the whole signal and the instantaneous descriptors that are computed for each time frame. The time frame is a short segmentation of the signal with a regular duration. In this paper, the duration for the time frame is 20 ms. As the proposed study focuses on the signal analysis of time frames, we adopt the instantaneous features as the acoustic characteristics for machine learning [10]. In 2004, G Peeters summarized a set of audio descriptors [11], including the temporal shape, temporal feature, energy features, spectral shape features, and perceptual features. This paper adopts Peeters' classification as the main method for extracting acoustic features to identify anomalous sounds. The descriptors for further machine learning processing include the following:

Temporal shape

Features (global or instantaneous) computed from the waveform or the signal energy (envelop). The attack time, temporal increase/decrease and effective duration are features of this category.

Frequency [12] Frequency is one of the basic features when describing or recognizing audio signals. In contrast to the time domain, which calculates the distance between two domain samples, the frequency is used to calculate the period vibration of two frequency band index bins. In this paper, short-time Fourier transform (STFT)-based analysis is applied for linear frequency calculations of continuous audio signals.

Amplitude The amplitude is a descriptor that represents the waveform shape with limited information. Similar to the processing steps of frequency, in this paper, the amplitude is calculated based on the continuous signals after the STFT and is converted to db-scaled from the logarithm scale.

Temporal features

Autocorrelation coefficients [11] The cross-correlation of a signal, as the inverse Fourier transform of the spectrum energy distribution of the signal, represents the signal spectral distribution in the time domain. This descriptor was proven by Brown in 1998 to be a valid description for classification. The formula is:

$$x_{corr}(k) = \frac{1}{x(0)^2} \sum_{n=0}^{N-k-1} x(n) \cdot x(n+k) \tag{1}$$

Each coefficient is in the range of [−1,1]. The faster the coefficients decrease with increasing lag, the whiter the signal can be.

Zero-crossing rate [12] The zero-crossing rate is a low-level feature used to describe the number of changes in signal values when crossing the zero axis. The concept assumes that the arithmetic mean of the audio signals is zero. The higher the zero-crossing rate is, the more high-frequency content there is, and the less periodic the audio signals are assumed to be.

Spectral shape features

Onset envelope Onset is the percept related to the time a sound takes to start. The onset envelope is computed as a spectral flux onset strength envelope. The spectral flux measures the amount of change in the spectral shape as the average difference between consecutive STFT frames. The onset strength at time t is determined by:

$$\sum_{m=1}^{m=M} H(X_{log,filt}(n,m) - X^{max}_{log,filt}(n-\mu,m)) \tag{2}$$

where ref is the logarithmically scaled filtered spectrogram $X_{log,filt}(n,m)$ after local max filtering $X^{max}_{log,filt}(n-\mu,m)$ along the frequency axis [13].

Onset is correlated with the logarithm of the attack time [14].

Spectral centroid [12] The spectral centroid represents the center of gravity (COG) of spectral energy. It is calculated as the frequency-weighted sum of the spectrum normalized by its unweighted sum:

$$v_{SC}(n) = \frac{\sum_{k=0}^{\frac{K}{2}} k \cdot |X(k,n)|^2}{\sum_{k=0}^{\frac{K}{2}} |X(k,n)|^2} \tag{3}$$

Spectral roll-off [12]

The spectral roll-off measures the bandwidth of the analyzed block n of the audio samples. The spectral roll-off point is the frequency at which the accumulated magnitudes of the STFT X (k, n) reach K of the overall sum of magnitudes:

$$v_{SR}(n) = k_r \left| \sum_{k=0}^{i} |X(k,n)| = K \cdot \sum_{k=0}^{\frac{K}{2}} |X(k,n)| \right. \tag{4}$$

The common value for K was 0.85 (85%). The spectral roll-off range is [0, K/2].

Mel-frequency cepstral coefficients (MFCC) [15]

The MFCC is defined as the compact description of the shape of the spectral envelope of an audio signal. It is calculated by the logarithm of the spectrum after the discrete cosine transform (DCT) or Fourier transform (e.g., FFT). Since MFCC was introduced in 1980, it has proven to be a valid measurement of audio signal classification to contain principal information. In our approach, the number of coefficients is 20.

Other features' categories

Intensity Intensity is a physical and measurable entity that is related to human perception of the magnitude of an audio signal. In this category, most features are instantaneous features, such as the root mean square and root mean square energy.

- Root mean squared energy. The RMS energy is calculated from the audio samples or from a spectrogram without

STFT processing. The advantage of the RMSE is the faster calculation speed because it does not require STFT processing. It outputs the RMS of each frame. In this paper, we only calculate the RMSE directly based on the audio signals.

Derived features

- Tempogram [16]. As a descriptor of the speed or pace of a given piece, a tempogram is usually measured in beats per minute (bpm). It is derived from the local autocorrelation of the onset strength envelope. For time t ϵ Z and time lag l ϵ [0, N]. W denotes window function: Z -> $\mathbb{R}$ centered at t = 0 with support [-N: N], N ϵ $\mathbb{N}$.

Enhanced Incremental Principal Component Analysis

Principal component analysis (PCA) is a classical multivariate statistical method for linear dimension reduction. It was introduced by Pearson as early as 1901 and Hotelling in the 1930s. As an unsupervised algorithm, the principal of PCA is to seek the subspace of the largest variance in the dataset. In 1982, the neural network implementation of one-dimensional PCA implemented by Hebb learning was introduced by Oja, and in 1989, it was expanded to hierarchical, multidimensional PCA by Sanger [17].

The enhanced incremental algorithm is based on the sequential Karhunen–Loeve (SKL) algorithm of Levy and Lindenbaum (2000) [18]. The computational advantages of the SKL algorithm are that it updates the original eigenspace and mean continuously with the learning rate, and the space complexity and the computational requirements are reduced to $O(d(k+m))$ and $O\left(dm^2\right)$, respectively, because it maintains constant space and time complexity in n. The disadvantage is that it does not calculate the varying sample mean of the training data with the new data. To resolve this issue, the enhanced incremental PCA is improved by adding an additional vector to the new training data to correct the time-varying mean [19].

In this paper, the input parameter of the enhanced IPCA, the number of components, is selected by a genetic algorithm based on the most optimized historical results of different machine types, which will be introduced in detail in "Automatic EPS Calculation (AEC)—Guided Genetic Algorithm".

Automatic EPS Calculation (AEC)—Guided Genetic Algorithm

The genetic algorithm (GA), a type of global stochastic search algorithm that includes evolutionary algorithms, particle swarm optimization and other biobased search methods, is applied for the selection of wrapper features [20]. Despite the capability of global searching, the exponentially increased computational cost of each candidate parameter restricts the efficiency of the GA. Therefore, the constraint of local optimization is added to resolve this issue.

Automatic EPS calculations (AECs) of randomly selected training datasets are used to set up the baseline of the initial range of estimated values of the candidate parameters. The wrapper parameters to be calculated in the guided genetic algorithm include the number of components for IPCA, the optimal epsilon value and the MinPts for DBSCAN. The automatic EPS calculation (AEC) algorithm estimates the EPS and MinPts based on the density of the randomly selected training datasets and the distances between the points in the density region. In the proposed AEC algorithm, the densities are calculated by the Gaussian kernel after the training dataset is scaled by MinMax. Similarly, the distances are calculated by the KD-Tree query after the MinMax scaled training dataset. The set of the estimated EPS and the estimated MinPts are the minimum values in all clusters. The range of the estimated number of components is set between 2 and 10 [21].

The three locally optimized parameters are input as the baseline to set up the range of values of the candidate parameters. The predicted value, the actual value, the difference between the predicted and the actual values, the mean squared error (MSE), the candidate number of components, the candidate EPS and the candidate MinPts, which are seven genes, are used to construct the chromosome. The fitness process is to set the reward value to 1 if the MSE is less than the target value of 0.4. Only the rewarded chromosomes construct the population for crossover and mutation to generate a new generation of populations with the preset crossover probability and specific mutation power [20].

Density-Based Spatial Clustering of Applications with Noise

DBSCAN was proposed by Martin Ester, Hans-Peter Kriegel, Jorg Sander and Xiaowei Xu in 1996. As a density-based clustering algorithm, DBSCAN separates clusters into low-density regions [22]. DBSCAN can identify global anomalies by defining dense and arbitrary shapes globally and, therefore, fails to identify local anomalies. There are two main advantages of DBSCAN over other unsupervised ML algorithms. The first is that DBSCAN does not require defining how many clusters to be calculated as an input parameter. It can define clusters of arbitrary shape by itself. Second, DBSCAN can handle noise points. With these two advantages, DBSCAN performs well when training and predicting large-volume and unbalanced datasets.

In DBSCAN, for any arbitrary object p belonging to dataset D, as shown in Fig. 1, the algorithm retrieves all object

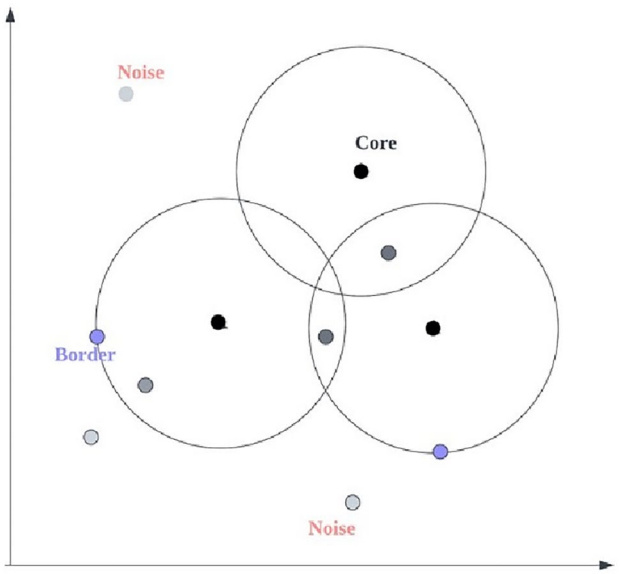

Fig. 1 Three scenarios of DBSCAN: core, border and noise points

densities reachable from p by the ε and MinPts values [22]. There are three scenarios for any object p: it is the core object of a cluster, if there are enough other objects q within the distance from $p \leq \varepsilon$ and with the count of $q \geq$ MinPts in dataset D; it is the border object if there is not enough q to be density-connected to p; it is the noise object if it does not belong to any cluster. The algorithm will continue processing to locate all the objects into clusters or noise groups.

In the hybrid algorithm proposed in this study, automatic EPS calculation (AEC) is adopted to estimate the EPS based on the average distance between the points of the training dataset, and MinPts is based on the kernel density of the training dataset, which includes the extracted acoustic features of the audio files. The assumption of the experiment is that the frames of the normal and anormal files have significantly different density characteristics so that they can be easily differentiated by the hybrid algorithm with reduced dimensions.

Experiments

Dataset and Preprocessing

The data were collected from machines in a plant in Suzhou City, China. The data consist of the normal/anomalous sounds of real machines. Each recording is a single-channel 2-s long audio of both a target machine's operating sound and environmental noise. The sample rate was 44,100. The audio files for the experiments can be downloaded via weblink (sharontan6217/asd (github.com)).

In the experiment, the training dataset includes unlabeled normal and anomalous datasets, in which 190 files are randomly selected from 228 normal audio files and 20 files from 120 abnormal data files, 50 consecutive times. The test dataset includes 20 unlabeled abnormal audio files. Ten acoustic features, e.g., frequency and amplitude, from the audio files are extracted as the components for clustering.

Benchmark System and Results

The benchmark performance of a deep convolutional neural network (DCGAN) is adopted for the experiment. The DCGAN is a deep convolutional neural network architecture composed of a pair of adversarial models called the generator and the discriminator [23, 24]. The generator creates a noise vector as the fake input of the discriminator. The discriminator segments the real and fake data distributions with certain policies. The details of the parameters are listed in Table 1.

Table 2 Experimental Results of the DCGAN shows the results of the benchmark experiments. The benchmark

Table 1 Parameters of the DCGAN

Hyperparameter	Setting
Optimizer	Admax of discriminator (lr = 1e−5, beta1 = 0.5), Admax of generator (lr = 1e−5, beta1 = 0.5), Admax of compiled (lr = 2e−5, beta1 = 0.5)
Dropout rate	0.1
Batch normalization	Momentum = 0.8
Leaky ReLU	Alpha = 0.2
Batch size	16
Noise initializer	Random uniform (−1,1)
Loss	Mean squared error
Monitor	Mean absolute error
Epoch	1200
GPU units	2
GPU cores	8

Table 2 Experimental results of the DCGAN

AUC	PAUC	F1 Score	MSE	Spearman rank correlation coefficient	Average execution time
0.77	0.69	0.59	0.48	0.47	90 min

algorithm of the DCGAN achieves an accuracy of 0.7. However, the average execution time of the DCGAN is 90 min with 2 GPU units. The computational cost of DCGANs is relatively high compared with that of machine learning algorithms.

Training Process

The architecture of the algorithm is to extract acoustic features from audio files collected in a real manufacturing environment. After the MinMax scaling, the normalized acoustic feature data are loaded into the layer of optimizations to calculate the parameters to construct the incremental principal analysis for dimension reduction and the DBSCAN clustering algorithm to detect the anomalous sound file (see Fig. 2).

During training, when optimizing the parameters via the AEC-guided genetic algorithm, the ranges for defining each parameter are based on the number of generations and the EPS and MinPS calculated via the AEC algorithm. The genetic algorithm selects the optimized parameters for which the compiled loss measure is less than the preset target value. Only the parameters selected by the guided genetic algorithm in the training are loaded to construct the dimension-reduction layer and the clustering algorithm to predict anomalous sound using the test dataset.

Algorithm AEC-guided genetic algorithm with IPCA-based DBSCAN

***Input**: Audio dataset; Training and testing, testing.*

***Output:** ASD results with forecast and AUC, PAUC, F1 Score, MSE, Hamming Distance, Jaccard Score, Spearman Rank Correlation Coefficient*

1. *Dataset = Audio dataset*
2. *Train = Train AEC-Guided Genetic Algorithm ("GA"), Incremental Principal Component Analysis ("IPCA"), Density-based Spatial Clustering of Applications with Noise ("DBSCAN").*
3. *analyzeAlgorithm (Genetic Algorithm).*
4. *TrainingOption:*
   crossover_prob = 0.6
   mutation_power = 0.4
   targetMSE=0.2
   num_generations = 100
5. *Load Audio Training Dataset and Test Dataset.*
6. *Extract Acoustic Features: Amplitude, Frequency, Autocorrelation, Zero crossing, RMSE, Pitch, Spectral Centroid, Spectral roll-off, Onset Envelop, Tempogram.*
7. *Calculate EPS and MinPts with the AES algorithm.*
8. *Set up ranges of the parameters for the optimization by Genetic Algorithm based on the calculated values in step 8.*
9. *Train Load Dataset:*
   With the training dataset, calculate forecasted labels with IPCA-based DBSCAN, and Mean Squared Error ("MSE") between actual and predicted labels.
   The parameters to be optimized (original inputs), the actual labels, the predicted labels, and the MSE construct the chromosome for the GA to process.
   GA selects the chromosome only when the gene of MSE is less than the target rate for the scaled datasets.
   The selected chromosomes continue for crossover and mutation to generate the new one.
   The newly generated outputs are the optimized parameters for the test dataset.
10. *Test Trained models/* using TestingData */.*
11. *Return TestResults = Labels of test datasets, validation matrix including AUC, PAUC, F1 Score, MSE, Hamming Distance, Jaccard Score, Spearman Rank Correlation Coefficient.*

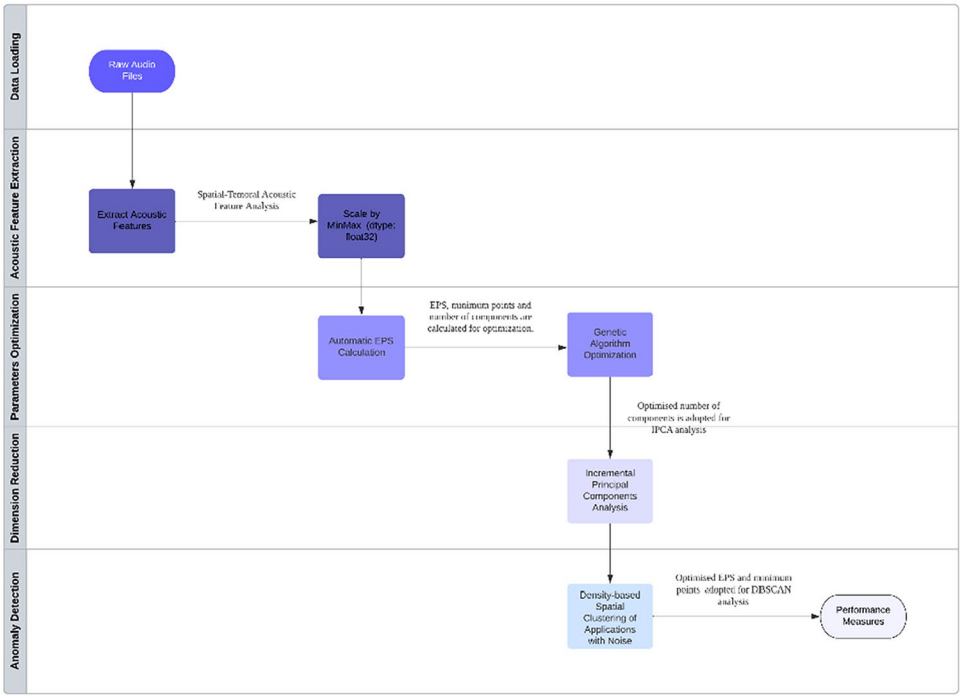

Fig. 2 Using self-adaptive IPCA-based DBSCAN to detect anomalous sound data

Results and Analysis

Table 3 shows a summary of the test results. According to Table 3, the performances of the AEC-guided GA and IPCA + DBSCAN are acceptable, with an average fitness of 0.843 and an average MSE of 0.16. The average execution time is less than 0.5 min for a total data size of 202,860,000 (training dataset: 185,220,000, test dataset: 17,640,000).

Figure 3 shows a sample of the prediction performance of the normal audio data changing to anomalous audio data. The normal class is set as "0", and the anomalous class is set as "1". The red line is the predicted clustering class, and the blue line is the actual class. The results show that the AEC-guided GA and IPCA-based DBSCAN models predict the turning point with high accuracy, and the AUC is 0.95.

Figure 4 shows the IPCA-based DBSCAN clustering results for a sample. With the optimized parameters, the normal and anomalous sound data are clearly clustered into two groups.

Table 4 shows the AUC, NMI, and F1 measure comparisons among 6 unsupervised and semisupervised machine learning or deep learning algorithms: K-means++, one-class SVM, agglomerative clustering, DCGAN, DCNN-Autoencoder, and AEC-Guided GA and IPCA + DBSCAN.

The experimental results show that both the AEC-guided genetic algorithm and IPCA-based DBSCAN for the extracted acoustic features and the DCNN-autoencoder for the audio data show the highest accuracy, with average AUCs of 0.843 and 0.8188, respectively. However, for the stability measures, the AEC-guided GA and IPCA-based DBSCAN of extracted acoustic features show the highest stability among all six semi-supervised or unsupervised algorithms [25], with the lowest Hamming loss of 0.16 and the highest Spearman rank correlation coefficient of 0.72.

Figure 5 shows the ROC curves of the six semisupervised and unsupervised machine learning algorithms. From the graph, it can be observed that the extracted acoustic features of the AEC-guided GA and IPCA-based DBSCAN algorithms reach the highest AUC value of 0.95, while the DCNN-AE and DCGAN algorithms achieve lower AUCs of 0.84 and 0.719864, respectively. The performances of agglomerative cluster and k-means++are the worst, at 0.65 and 0.60, respectively.

Noise Tolerance Test

Another series of experiments is conducted to test the maximum noise tolerance of the AEC-guided genetic algorithm and

Table 3 Summary of performance evaluation

Number of test cases	AUC	PAUC	MSE		Spearman rank correlation coefficient	Average execution time
50	0.84	0.58	0.16	0.72		0.5 min

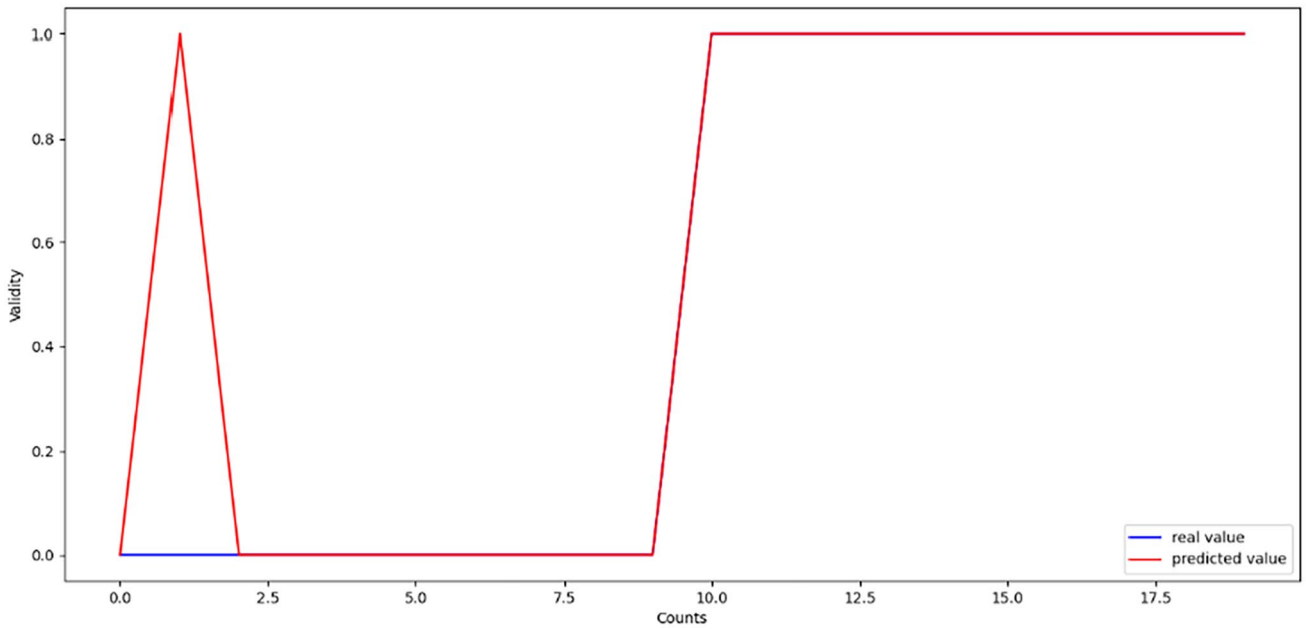

Fig. 3 ROC curves of IPCA-based DBSCAN algorithms

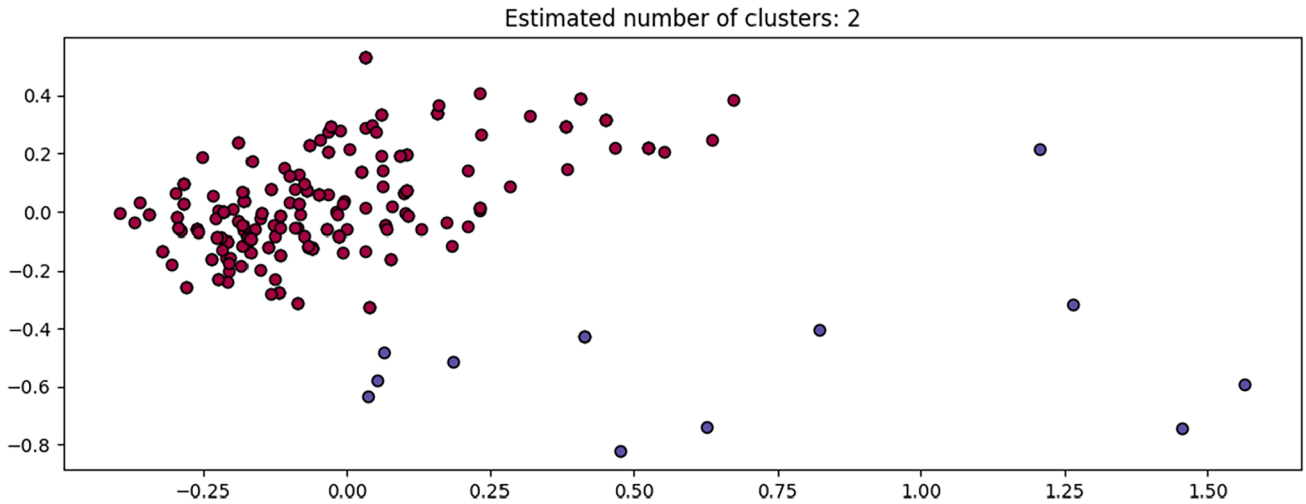

Fig. 4 Clustering of IPCA-based DBSCAN algorithms

Table 4 AUC Comparison between unsupervised and semisupervised Ml algorithms

Machine learning	AUC	F1 score	MSE	Hamming distance	Spearman rank correlation coefficient
K-means++	0.54	0.54	0.47	0.47	0.071
One-class SVM	0.73	0.73	0.27	0.27	0.55
Aggregate clustering	0.58	0.58	0.42	0.42	0.13
DCGAN	0.77	0.6	0.41	0.41	0.47
DCNN-autoencoder	0.82	0.5	0.42	0.5	0.55
AEC-guided GA and IPCA + DBSCAN	0.84	0.84	0.16	0.16	0.72

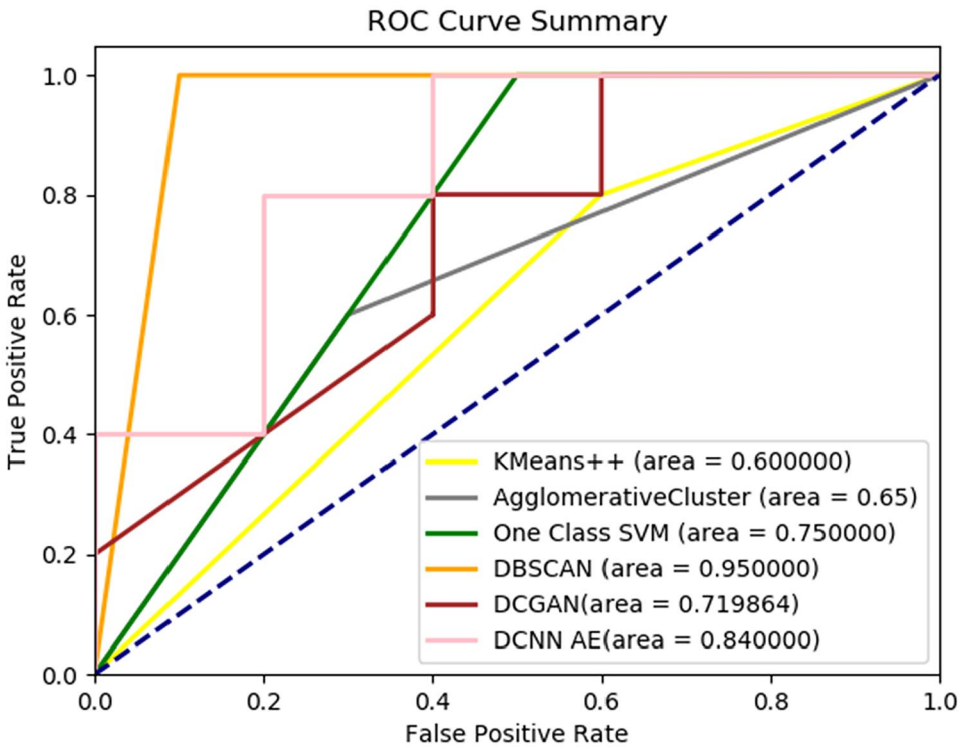

Fig. 5 ROC curves of the unsupervised and semisupervised ML/DL algorithms

Table 5 Comparison between the hardcoded architecture and the parameterized architecture with 50 random test cases

Machine learning	AUC	PAUC	F1 score	MSE	Jaccard score
Hardcoded architecture	0.82	0.57	0.82	0.18	0.72
Automatic EPS calculation	0.84	0.58	0.84	0.16	0.76

IPCA-based DBSCAN. Based on the experimental results, the performance of the algorithm is impacted when the SNR is 13.0103 (SNR = 10*$\log_{10}$(1/0.05)), in which 0.05 is the noise significant factor [26]. This is because DBSCAN is unable to detect and filter noise outliers instead of the continuous noise pattern added to the clean audio sample. This is the disadvantage observed when it is applied to lab experiments.

Comparison of the Hardcoded Architecture and Parameterized Architecture

In the experiments to compare the hardcoded architecture and the parameterized architecture, it is observed that the parameterized architecture requires less execution time and achieves high accuracy. In this experiment, the hardcoded architecture is set to an EPS of 0.07, and the MinPts is set to 2. The experimental results of 50 random test cases shown in Table 5 indicate that although the AUC of the hardcoded architecture is 0.82, the stability indicators, including the Jaccard score and Spearman rank correlation coefficient, are significantly lower than those of the parameterized architecture. Therefore, the performance of the hardcoded architecture is not so satisfactory as that of the parameterized architecture.

Conclusions and Future Work

The hybrid algorithm to integrate the AEC-guided genetic algorithm and IPCA with DBSCAN for anomaly sound detection seems to be a promising direction for ASD when handling different environmental issues and different types of audio files. Notably, when detecting rare events in multiple scenes (including silence and background sounds), the proposed unsupervised algorithm did not perform as well as the machine sounds. This is possibly due to the quality of the collected sound because we used high-quality equipment to collect the machines' sounds at the specific plant site. Future research will improve the noise tolerance of the algorithm for environments with mixed sounds.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42979-024-02844-y.

Author Contributions Xiao Tan: Primary Contact, Siu-Ming Yiu: Second Author Contribution.

Funding Not applicable.

Data Availability Audio data is collected from real plant-sites. The data files for experiments can be referenced via the weblink of GitHub: sharontan6217/asd (github.com).

Declarations

Conflict of interest Not applicable.

Informed Consent Not applicable.

Research Involving Human and/or Animals Not applicable.

References

1. Ismail B, et al. Industrial IoT. Springer International Publishing. 2020.
2. Ajagbe SA, et al. Performance investigation of two-stage detection techniques using traffic light detection dataset. IAES Int J Artif Intell (IJ-AI). 2023;12(4):1909–19.
3. Diaz C, et al. Traffic flow indicators analysis to determine causes of vehicular congestion. ParadigmPlus. 2021;2(2):1–16.
4. Rabelo L, et al. Using delphi and system dynamics to study the cybersecurity of the IoT-based smart grids. ParadigmPlus. 2022;3(1):19–36.
5. Rawat R, et al. Autonomous artificial intelligence systems for fraud detection and forensics in dark web environments. Informatica 47.9. 2023. https://doi.org/10.31449/inf.v46i9.4538.
6. Grueneberg K, Ko B, Wood D, Wang X, Steuer D, Purohit YL. IoT data management system for rapid development of machine learning models In: IEEE, International Conference on cognitive computing (ICCC); 2019.
7. Kuncheva LI. A stability index for feature selection artificial intelligence and applications. 2007;421–7.
8. Hisashi U, Yuma K, Shoichiro S, Akira N, Noboru H. Anomaly detection technique in sound to detect faulty equipment. NTT Tech Rev. 2017;15.8 (2017):28–34. https://doi.org/10.53829/ntr201708fa5.
9. Gorawski M, Malczok R. AEC algorithm: a heuristic approach to calculating density-based clustering Eps parameter. In: Advances in Information Systems: 4th International Conference, ADVIS 2006, Izmir, Turkey, Proceedings 4. Berlin Heidelberg: Springer; 2006.
10. Heittola TC, Virtanen E, Marcin T. The machine learning approach for analysis of sound scenes and events. In: Virtanen T, Plumbley M, Ellis D, editors. Computational analysis of sound scenes and events. Cham: Springer; 2017. https://doi.org/10.1007/978-3-319-63450-02.
11. Peeters G. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. 1st CUIDADO Project Report. 2004;54:1–25.
12. Lerch A. An introduction to audio content analysis: applications in signal processing and music informatics. Wiley-IEEE Press; 2012.
13. Böck S, Widmer G. Maximum filter vibrato suppression for onset detection. In: 16th International Conference on digital audio effects, Maynooth, Ireland. 2013.
14. Lemaitre G, Grimault N, Suied C. Acoustics and psychoacoustics of sound scenes and events. Comput Anal Sound Scenes Events. 2018;41–67. https://doi.org/10.1007/978-3-319-63450-0_3.
15. Abdul ZK, Al-Talabani AK. Mel frequency cepstral coefficient and its applications: a review. IEEE Access. 2022;10:122136–58.
16. Grosche P, Müller M, Kurth F: Cyclic tempogram—a mid-level tempo representation for musicsignals. 2010 IEEE International Conference on acoustics, speech and signal processing. IEEE; 2010.
17. Mishra SP, et al. Multivariate statistical data analysis-principal component analysis (PCA). Int J Livest Res. 2017;7(5):60–78.
18. Ross DA, et al. Incremental learning for robust visual tracking. Int J Comput Vis. 2008;77:125–41.
19. Wang J, et al. A fast incremental multilinear principal component analysis algorithm. Int J Innov Comput Inf Control. 2011;7:6019–40.
20. Tan X. Libor prediction using genetic algorithm and genetic algorithm integrated with recurrent neural network. In: 2019 Global Conference for advancement in technology (GCAT). IEEE; 2019.
21. Gorawski M, Malczok R. Towards automatic Eps calculation in density-based clustering. Advances in databases and information systems: 10th east european conference, ADBIS 2006, Thessaloniki, Greece, September 3–7, 2006 proceedings 10. Springer, Berlin, Heidelberg, 2006.
22. Parimala M, Lopez D, Senthilkumar NC. A survey on density based clustering algorithms for mining large spatial databases. Int J Adv Sci Technol. 2011;31(1):59–66.
23. Lee YO, Jo J, Hwang J. Application of deep neural network and generative adversarial network to industrial maintenance: a case study of induction motor fault detection. In: 2017 IEEE, International Conference on big data (BIGDATA).
24. Kopčan J, Škvarek O, Klimo M. Anomaly detection using autoencoders and deep convolution generative adversarial networks. In: 14th International scientific Conference on sustainable, modern and safe transport.
25. Khoshgoftaar TM, et al. A survey of stability analysis of feature subset selection techniques. In: 2013 IEEE, 14th International Conference on information reuse & integration (IRI). IEEE, 2013.
26. Plapous C, Marro C, Scalart P. Improved signal-to-noise ratio estimation for speech enhancement. IEEE Trans Audio, Speech, Lang Process. 2006;14(6):2098–108.