Visual Informatics 6 (2022) 47–55

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

AFExplorer: Visual analysis and interactive selection of audio features

Lei Wang, Guodao Sun *, Yunchao Wang, Ji Ma, Xiaomin Zhao, Ronghua Liang

College of Computer Science, Zhejiang University of Technology, Hangzhou, 310023, China

ARTICLE INFO

Article history:
Received 12 September 2021
Received in revised form 14 December 2021
Accepted 20 February 2022
Available online 4 March 2022

Keywords:
Audio data
Interactive feature selection
Visual analytics
Visualization systems and tools

ABSTRACT

Acoustic quality detection is vital in the manufactured products quality control field since it represents the conditions of machines or products. Recent work employed machine learning models in manufactured audio data to detect anomalous patterns. A major challenge is how to select applicable audio features to meliorate model's accuracy and precision. To relax this challenge, we extract and analyze three audio feature types including Time Domain Feature, Frequency Domain Feature, and Cepstrum Feature to help identify the potential linear and non-linear relationships. In addition, we design a visual analysis system, namely AFExplorer, to assist data scientists in extracting audio features and selecting potential feature combinations. AFExplorer integrates four main views to present detailed distribution and relevance of the audio features, which helps users observe the impact of features visually in the feature selection. We perform the case study with AFExplore according to the ToyADMOS and MIMII Dataset to demonstrate the usability and effectiveness of the proposed system.

1. Introduction

Acoustic quality detection in manufactured products quality control is to detect the quality of the product based on the audio data which is generated when the product is running. In recent years, data scientists and researchers employed machine learning models in manufactured acoustic quality detection to improve the accuracy and efficiency of detection while aiming to reduce labor costs.

Machine learning is proven to be an effective technique in numerous fields such as the biomedicine (Palaniappan et al., 2013), manufacturing industry (Zeng et al., 2009), and speech recognition (Wang et al., 2018). However, the performance of the machine learning model depends on the feature selection in different scenes, which swamps almost machine learning novices and senior researchers. Particularly in the manufactured acoustic quality detection, manufacture audio contains complex features such as frequency feature, temporal feature, and noise chaos. The features used as the input of machine learning models are either Mel Frequency Cepstral Coefficients (MFCC) or Linear Predictive Cepstral Coefficients (LPCC) (Fang et al., 2001; Sousa et al., 2019). Both MFCC and LPCC are presented by approximate (12–40) components (Zhou et al., 2011). Components selection of audio features without further analysis could lead to the lack of audio information such as the components in MFCC cannot represent high-frequency information of audio. Therefore, how to select and evaluate feature components to represent manufactured audio information is a vital challenge when performing manufactured acoustic quality detection.

The features of data could represent the data information in a more efficient and brief form. However, it is improper to use the whole data feature in the practiced scene. For example, in audio quality detection of manufactured products, we cannot employ every audio feature due to the audio data contains many interference factors such as the features of noises. Feature engineering can help us to solve this problem by generating, extracting, deleting, and combining features of data. Feature engineering is a process that converts raw data into a set of features with a better expression of an underlying problem. Feature engineering is used in many fields (e.g. image recognition, text classification, and audio data feature extraction) due to its ability to make machine learning models with more predictive and comprehensive performance in certain scenes (Khurana et al., 2017). In recent years, automated and semi-automated feature engineering methods are proposed to help users select features to convey data information, while both kinds of methods need a piece of strong domain knowledge to evaluate the performance of selected features. How to compare different feature selection methods on audio data is the other challenge that is confronted by data scientists and researchers.

In this paper, we propose a novel method focusing on relaxing the below two challenges. *One challenge is how to select and evaluate audio features and feature components subset in audio quality detection of manufacture products*, and *the other challenge is how to compare different feature selection methods on audio*

* Corresponding author.
E-mail address: guodao@zjut.edu.cn (G. Sun).

https://doi.org/10.1016/j.visinf.2022.02.003

data. Our method integrates three main feature selection methods including filtering method, wrapper method, and embedded method (Chandrashekar and Sahin, 2014; Guyon and Elisseeff, 2003). Our technique allows users to alter filter conditions and thresholds interactively to investigate feature relationships. In addition, our method supports users to modify feature subset (e.g. adding or removing a feature) to explore the performance of selected features. We perform feature selection methods to calculate the importance of all the features and present the rank of features by their importance value to assist users in selecting important features. To help users select and evaluate features, we employ a multi-view linked digging technique and design an audio features exploration visual analysis system named AFExplorer. In summary, the primary contributions of this paper are as follows:

- We propose an interactive method integrating three feature selection methods including filtering method, wrapper method, and embedded method to help data scientists and researchers select and evaluate the audio feature in audio quality detection of manufactured products.
- We design and implement a visual analysis system (AFExplorer) that combines multi-view to assist users in exploring and comparing the performance of interactively selected audio features in audio quality detection.
- We present the case study and conduct a comprehensive evaluation to demonstrate the effectiveness of the AFExplorer.

The rest of the paper is organized as follows: Section 2 summarizes the related work; Section 3 introduces system workflow and the design goals for the challenges in audio data; Section 4 demonstrates the interactive system construction process; Section 5 presents the case study we conducted; Section 6 presents the evaluation from domain experts, and Section 7 gives our conclusions.

2. Related work

In this section, we elaborate on the related work according to the following three aspects: (1) Audio feature Extraction and Generation; (2) Feature Selection Method and (3) Visual Analytics in Feature Selection.

2.1. Audio feature extraction and generation

Audio feature extraction plays an important role in the field of audio processing, hence it has attracted the attention of many researchers. There are some efficient toolboxes (e.g., MIR Toolbox, Essentia, Libxtract, and Aubio) developed to extract features from audio data (Brossier, 2006; Bullock, 2007; Bogdanov et al., 2013; Lartillot and Toiviainen, 2007). These tools promote work and studies progress with their peculiarity on the parts of real-time and high-dimensional feature extraction. Furthermore, Moffat et al. (2015) evaluated existing audio feature extraction libraries in terms of coverage, effort, presentation, and time lag.

In addition, other work dedicated to audio feature generation from exited audio features by algorithms and models (Mierswa and Morik, 2005; Hamel and Eck, 2010). Hamel and Eck (2010) applied Deep Belief Network (DBN) on Discrete Fourier Transforms (DFTs) of the audio to extract features. Comparing with MFCCs, the learned features perform significantly better. They (Liu et al., 2021) proposed the architecture that took the multi-scale time-series information into consideration, which transfers more suitable semantic features for the decision-making. Janssens et al. (2016) proposed a feature learning model based on convolutional neural networks (CNN) for condition monitoring, which utilized manually engineered features and a random forest classifier and performed better than the traditional method.

2.2. Feature selection method

Feature selection is a common preprocessing step in machine learning algorithms. Feature selection has good performance in reducing the redundancy of data, removing irrelevant data, improving the accuracy of learning models, reducing computational complexity, and improving the explainability and interpretability of model results. Feature selection can be divided into three categories: (a) filter methods, (b) wrapper methods, and (c) embedded methods.

In filter methods, various evaluation indicators (e.g., correlation measurement, distance measurement, information measurement Lewis, 1992, and consistency measurement Sun et al., 2013) are used to measure the correlation between each function and category. Variable ranking techniques are utilized as the principal criteria for variable selection, and a threshold is set to filter out the less relevant variables. These methods improve the classification accuracy of most classifiers and reduce computational complexity. Especially when dealing with large-scale data or online data, these advantages are more obvious. Currently, certain entropy-based feature selection models have been proposed (Battiti, 1994; Hancer et al., 2018). Lewis (1992) proposed the most basic information gain theory based on mutual information. Based on this basis, Battiti's mutual information feature selection (mutual information for selecting features, MIFS) method and Peng's maximum correlation minimum redundancy (mRMR) method enriched the mutual information theory (Sun et al., 2013; Sindhwani et al., 2004).

Wrapper methods typically use the evaluation function to implicitly select the features and evaluate them through the model features returned by the learning algorithm. Selecting subsets of features according to the accuracy of classification algorithms, which can be regarded as black boxes that score subsets of features. The wrapped method is slower than the filter method in speed. However, the size of the optimized feature subset computed by the wrapped method is relatively small, which is very conducive to the identification of key features. At the same time, its accuracy is relatively high, though its generalization ability is relatively poor. Yang et al. (2011) used decision trees for feature selection, and genetic algorithms were used to find a set of feature subsets by minimizing the classification error rate of decision trees. Chiang and Pell (2004) combined fisher discriminant analysis with the genetic algorithm to identify key variables in the process of chemical faults and achieved a better result. Guyon and Elisseeff (2003) used the classification performance of support vector machines to measure the importance of features and finally constructed a classifier with higher classification performance. Michalak and Kwaśnicka (2006) proposed a wrapper feature selection method based on a dual strategy of mutual relationships.

Embedded methods incorporate the feature selection as part of the training process and evaluate each feature set with the trained classifier. Predictably, wrapper methods take large feature space and computing time, which decelerates the process of feature selection. Computing time can be reduced by combining the filtering method and the packaging method. However, the embedded method mostly focuses on searching in a local space, and the coverage is limited.

In general, these technologies have feature extraction capabilities. Whereas these methods also face a common flaw that when facing this kind of black box system, data scientists and researchers have rare knowledge of why the features are selected. If the performance of features allows users to directly perceive, making users directly observe their connection and significance. The perception and interaction functions are exactly what the visual analysis method is good at. The AFExplore has designed feature analysis views according to various visual levels, enabling users to rapidly perceive the relevance and necessity of the audio features. It also provides users with interactive features to assist them in choosing the appropriate set of audio features.

2.3. Visual analytics in feature selection

There are various visualization techniques related to feature selection. For example, correlation matrices (Friendly, 2002; MacEachren et al., 2003), radial visualizations (Artur and Minghim, 2019; Sánchez et al., 2018; Turkay et al., 2011), scatterplot matrices (Wang et al., 2017), feature ranking (Elmqvist et al., 2008; Gratzl et al., 2013), feature clustering (Johansson and Johansson, 2009), and dimensionality reduction (DR) (Turkay et al., 2011; Sun et al., 2021). Guo (2003) introduced the idea of visualizing relationships between different feature sets. They proposed an interactive matrix view where rows and columns represent features and the cells are colored according to feature similarity. The automatically sorted matrix is helpful to select subspaces (i.e., feature subsets) where show interesting clusters. Muhlbacher and Piringer (2013) presented an interactive framework that displays feature ranking for building regression models, which aids users to understand the relationship between attributes and a target one. Krause et al. (2014) proposed a visual analysis tool called INFUSE, which enables the ensemble of multiple feature selection methods by visualizing features importance determined by various different feature selection methods in a radial glyph. Zhao et al. (2019) developed a visualization system called FeatureExplorer, which allows users to iteratively refine and diagnose the model by selecting features based on their domain knowledge.

However, the previous work mainly considered the importance of a single feature but ignored the correlation among different features and their impact on the importance of features. In order to reveal the relationship between the features in a more complete manner, the strength of linear and non-linear correlations is considered in the system. Three primary technologies of feature selection are also comprehensively is used and the corresponding view designed to facilitate users to compare the importance of each feature. In the process of feature analysis, we comprehensively consider the influence of features and design some interactions to help explore and analyze features. In addition, users are allowed to obtain prominent features, modify the size of the subsets, and obtain new feature subsets to compare with other subsets.

3. Task analysis and workflow

Following our above-mentioned literature review, we describe user tasks in this part. To help users explore and analyze audio features, we summarize four core analysis tasks that appear most frequently in the domain research.

T1 **Comparison of feature selection methods.** It is indispensable to compare three feature selection techniques we applied in the system to find different highlights. We can compute feature importance which reflects the results from feature selection techniques through diverse algorithms. These results can be intuitively perceived by users and facilitate the visual comparison of feature selection techniques for each feature. Hence, it is essential to provide appropriate interaction which allows users to select the feature selection technique they desiderate.

T2 **Exploration of feature correlation.** The system should provide users with detailed information on the different features. For example, the Pearson Correlation Coefficient is generally used to measure the linear relationship between different types of features. The Distance Correlation Coefficient is also utilized to assist users in perceiving the possible non-linear relationship between features. The correlation between features and categories is presented through Variance Influence Factor (O'brien, 2007). In terms of the visual design, the system should display detailed information about features as much as possible.

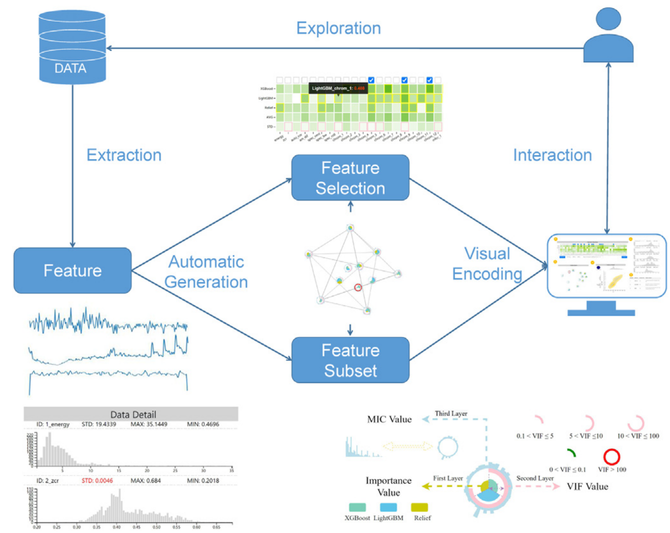

Fig. 1. The components diagram of AFExplorer.

T3 **Detection of feature performance.** Through T2, users have ability to find the specifics and significances of features. In addition, our system should provide certain interaction components and options to assist users in the process of analyzing. Users can further analyze the performance of the features through mouse events, interactive controls and interface interactions. Allow users to interactively select important features they are interested in.

T4 **Extraction of feature subset.** Our system utilizes an automatically generation approach to obtain feature subset. However, there is no guarantee about the quality of these feature subsets. Hence, it is convenient for users to select feature subset by combining automatic generation and interaction.

As shown in Fig. 1, the workflow of AFExplorer mainly includes three steps: (1) Feature Extraction; (2) Feature Selection; and (3) Visual Analysis.

In the Feature Extraction step, the system comprehensively extracts the information contained in each audio data as far as possible. Three most relevant features are utilized, including Time Domain Feature, Frequency Domain Feature, and Cepstrum Feature.

In the Feature Selection step, AFExplorer utilizes three feature selection algorithms to obtain the degree of importance of audio features. Quantification of audio features is utilized to determine the best feature subset generated by different algorithms.

In the Visual Analysis step, we design multiple views to analyze the feature importance and relation. Users can select multiple features with their demands to generate feature subset during the analysis process.

4. Data description

4.1. Audio feature extraction

We parse and preprocess the audio data from the ToyADMOS (Koizumi et al., 2019) and MIMII Dataset (Purohit et al.,

Table 1
Feature sets.

Feature type	Feature name	Dimension	Feature ID
Time domain feature	Short-term energy	1	1
	Zero crossing rate	1	2
	Auto correlation	1	3
	Average magnitude difference	1	4
Frequency domain feature	Spectral centroid	1	5
	Spectral bandwidth	1	6
	Spectral rolloff	1	7
	Chromagram	12	8–19
Cepstrum feature	Mel frequency cepstrum coefficient	12	20–31
	Linear prediction cepstrum coefficient	12	32–43

2019) to prepare for the next step of extracting features. The data comprises parts of ToyADMOS and the MIMII Dataset consisting of the normal/anomalous operating sounds of six types of toy/real machines. Each recording is a single-channel (approximately) 10-sec length audio that includes both a target machine's operating sound and environmental noise. The following three types of toy/real machines are used in our study: Toy-car (ToyADMOS), Valve (MIMII Dataset) and Fan (MIMII Dataset). Each machine's dataset consists of (i) around 3000 samples of normal sounds for training and (ii) 1600 samples each of normal and anomalous sounds for the test. There are three major methods, including time domain method, frequency domain method, and cepstrum analysis method for feature extraction of audio data.

Time Domain Feature. The time domain method mainly considers the waveform information of the audio signal. They reflect inherent properties over time, including the amplitude and frequency of changes in the audio signal. In the case of loud noise, the audio waveform will fluctuate particularly sharply. The Short-term Energy and Zero Crossing Rate can exhibit pronounced extreme peaks or troughs and irregular fluctuations in the time series. On the contrary, the Short-term Energy and Zero Crossing Rate will perform relatively smoothly without extreme spikes in a silent environment. Typical time domain features consist of Short-term Energy, Zero Crossing Rate, Auto Correlation, etc. (Sharma et al., 2020; Peeters, 2004).

Frequency Domain Feature. The frequency domain method is to analyze the frequency spectrum of the audio signal to obtain part of meaningful audio features, such as formant and bandwidth. It is also an essential method of audio signal processing. Frequency domain features include the Short-time Fourier Transform, Spectral Centroid, Spectral Bandwidth, etc. (Sharma et al., 2020; Peeters, 2004; Agostini et al., 2003).

Cepstrum Feature. The cepstrum features of audio signals include LPCC and MFCC. The signals are in many objective physical phenomena and the combination of its components is the product combination signal or the convolution combination signal. However, it is tough to analyze this kind of nonlinear system/problem. Conducting homomorphic analysis to convert it into a linear problem is indispensable. After homomorphic analysis of the audio signal, the cepstrum parameters of the audio signal will be obtained (Randall, 2017). Therefore, homomorphic analysis is also known as inverse cepstrum analysis. The specific audio features we extracted are shown in Table 1 (see Fig. 2).

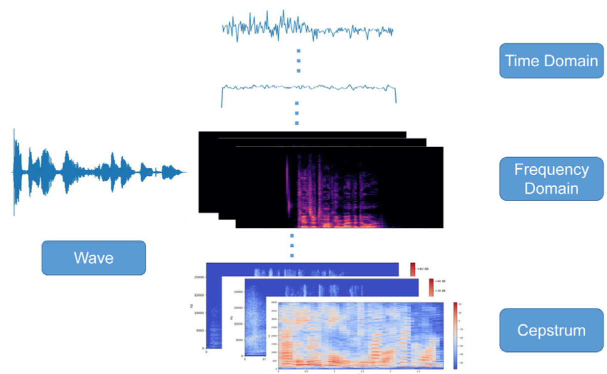

Fig. 2. Three types of audio features.

4.2. Feature selection algorithms

In Section 2, we introduced three types of feature selection methods, among which we chose the Relief algorithm (Kira and Rendell, 1992) in the filter method and the two wrapper methods including eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine(LightGBM) (Ke et al., 2017; Chen and Guestrin, 2016). Utilizing various classification algorithms as weak classifiers and with excellent usage of weak classifiers for cascading, Adaptive Boosting (Adaboost) possesses high accuracy. Inspired by the embedded method, we combined the REF method with AdaBoost to extract the audio feature subsets.

Since the Relief algorithm is simply constructed with high operating efficiency, it has been widely adopted by researchers. It is a feature weighting algorithm that assigns different weights to features according to the correlation of each feature and category. Features with a weight less than a certain threshold will be removed. The other two algorithms, XGoost and LightGBM belong to the Boosting method. The basic idea is to train the newly added weak classifier according to the current model loss function's negative gradient information and combine the trained weak classifier into the existing model in an accumulated form. In addition, LightGBM is an improvement to XGBoost which includes parallel schemes and gradient-based unilateral detection. The recursive feature elimination method employs a machine learning model (such as Support Vector Machine and Regression Models) for multiple rounds of training. Afterward, each round of training eliminates the features corresponding to some weight coefficients. The next round of training is performed based on the new feature set. We apply the above three algorithms to obtain the importance of each feature. AdaBoost gradually discards the least important audio features derived from the current method throughout training, until only the final (and most crucial) feature remains.

4.3. Relation indicators

Linear strength is often used for feature selection in machine learning. We calculate the correlation coefficient between different feature items, judge the strength of the correlation between the feature items according to the correlation coefficient and determine the importance of the feature items. Three correlation coefficients are used in this paper, including the Pearson Correlation Coefficient (PCC), Distance Correlation Coefficient (DCC), and Maximal Information Coefficient (MIC) (Moffat et al., 2015).

The Pearson Correlation Coefficient is a linear correlation coefficient and a statistic used to reflect the degree of linear correlation between two variables. The correlation coefficient is represented by r, which describes the strength of the linear correlation between two variables. The value of r is between -1 and $+1$. If $r > 0$, it indicates that the two variables are positively correlated. If $r < 0$, it means that the two variables are negatively correlated. So the larger the absolute value of r, the correlation becomes stronger.

The Distance Correlation Coefficient is created to overcome the weakness of the Pearson Correlation Coefficient. In the case of x and x^2, even if the Pearson Correlation Coefficient is 0, we cannot conclude that the two variables are independent. They may be possibly non-linearly related. If the Distance Correlation Coefficient is 0, we say that these two variables are independent.

$$\widehat{dcorr}(X,Y)=\frac{\widehat{dcov}(X,Y)}{\sqrt{\widehat{dcov}(X,X)\widehat{dcov}(Y,Y)}} \tag{1}$$

The Distance Correlation Coefficient is calculated using Eqs. (1)–(4). In the formula, $dcov^2(X,Y)=\widehat{S}_1+\widehat{S}_2-2\widehat{S}_3$, $\widehat{S}_1$, $\widehat{S}_2$, and $\widehat{S}_3$ are described as follows:

$$\widehat{S}_1=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|X_i-X_j\|d_X\|Y_i-Y_j\|d_Y \tag{2}$$

$$\widehat{S}_2=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|X_i-X_j\|d_X\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|Y_i-Y_j\|d_Y \tag{3}$$

$$\widehat{S}_3=\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\|X_i-X_l\|d_X\|Y_j-Y_l\|d_Y \tag{4}$$

MIC belongs to the method of Maximal Information-based Nonparametric Exploration (MINE) (Reshef et al., 2011). It is used to measure the degree of correlation between two variables X and Y, which is the strength of linearity or non-linearity. The MIC overcomes the shortcomings of Mutual information (MI). The MI method processes laboriously the continuous variables and its result is not capable to be measured and normalized. MIC utilizes an optimal discretization method to convert the value of mutual information into a measurement method, with the value interval in [0, 1]. The relationship and function between features can be analyzed in detail through the three coefficients. The MIC is calculated using Eqs. (5)–(6).

$$I[x;y]\approx I[X;Y]=\sum_{X,Y}p(X,Y)\log_2\frac{p(X,Y)}{p(X)p(Y)} \tag{5}$$

$$MIC[x;y]=\max_{|X||Y|<B}\frac{I[X;Y]}{\log_2(\min(|X|,|Y|))} \tag{6}$$

Variance Inflation Factor (VIF) measures how much the behavior (variance) of an independent feature is influenced, or inflated, by its correlation with the other independent feature. VIF allows a quick measure of how much a feature is contributing to the error. When $0 < VIF \le 5$, there is hardly any remaining collinearity. When $5 < VIF \le 10$, there is slight collinearity. When $10 < VIF \le 100$, there is strong collinearity. When $VIF \ge 100$, there is severe complex collinearity. If the VIF is too large, it means that there is a strong correlation between the independent features. Users can remove the feature with the larger VIF or combine the related features into a single feature.

5. Visual design

Following the analytical tasks, we developed AFExplorer, an interactive web-based VA system that approves users to utilize multiple metrics and feature selection algorithms in order to explore worthwhile features. The frontend is implemented in JavaScript using D3.js, the backend is written in Python with the Librosa and Sklearn packages. The system includes five visualization panels. (b) Feature Selection (T1, T4), the heatmap is designed to compare feature selection algorithms. (c) Feature Space (T2, T3), the glyph reveals correlation of features and interaction is designed to help users explore the correlation and difference of features. (d) Feature Detail (T2, T3), the scatter plot represents the detail of audio features. (a) Data Detail (T3) and (e) Feature Subset(T4), the Data Detail panel furnishes the detailed statistical index of features and the table (Feature Subset) panel furnishes subsets users selected.

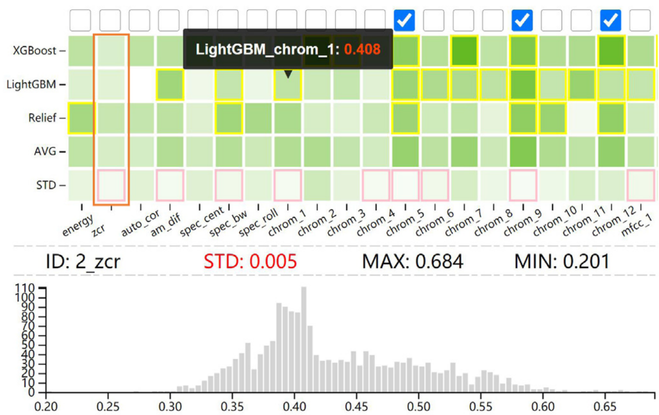

Fig. 3. Comparison of feature selection methods.

5.1. Feature selection

In the feature selection view, the heat map shows the score of each feature by the three automated feature selection methods. The average and variance of the importance score of each feature are added into the heat map, which assists users to visually perceive the different feature importance generated by different methods. For example, as shown in Fig. 3, the importance scores of feature Zero Crossing Rate (zcr) under the three methods are extremely low and represent inferior visibility. The feature zcr performs prominently in the variance column. In this view, a border of prominent color is added to the color block to show the best feature subset adopted by the current method. It helps users to identify representative feature sets from each technology, perceive the differences between different technologies through these elements, and select the feature subset they demand for a more detailed understanding in the Feature Space panel.

5.2. Glyph design

The first layer is designed for comparing the feature importance index of different feature selection algorithms. The opacity is filled by the average of these three indexes. How to compare the size of the three values at the same time? Here, we utilize the pie chart, each algorithm takes up one-third of the circle and different colors correspond to different algorithms. The length of the outer radius represents the importance values generated by different algorithms to the power of one-half. Additionally, in order to perceive the importance of this feature for users, we also add opacity to the visual design of glyph. The larger values of the three importance index of a feature, the more distinct the glyph presents.

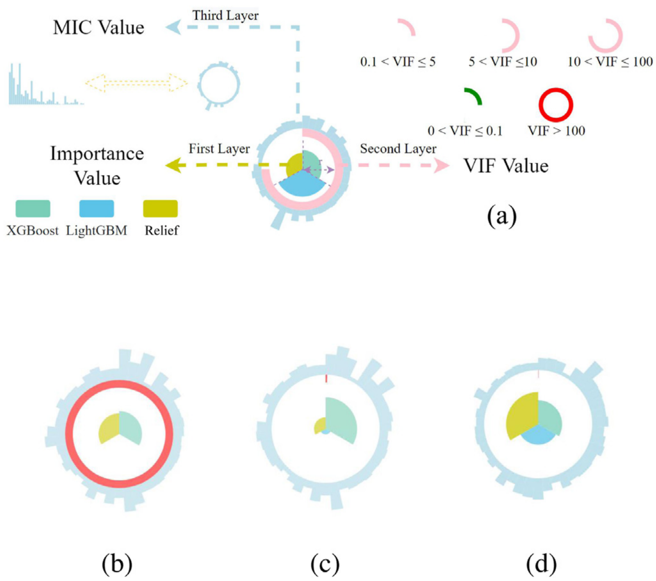

Fig. 4. Glyph design. (a) is the overall design of glyph. Since the range of VIF values are large, if the normalized VIF values are directly mapped to the arc length, it will lead to visual misinterpretation. The VIF value of features are invisibility in glyph. (b), (c), and (d) are the typical cases. Therefore, we apply the segmented representation method as shown in (a). When the maximum threshold is exceeded, the arc transforms into the red. When VIF value is less than the minimum threshold, the arc transforms into the green.

The second layer and the third layer of the glyph are to analyze influencing factors and comparability of features. Explicitly, the second layer analyzes from the feature itself, and the third layer sketches the MIC value among each feature. The second layer describes the VIF of each feature by arcs. We adopt the segmented display based on the variance inflation factor VIF value threshold. While normalizing the VIF value in the step of visual design, we find that some VIF values reveal singular values. We consider two extreme circumstances in visual design. As the VIF value exceeds the maximum threshold ($VIF > 100$), the entire arc will be colored red. If the VIF value is inferior to the minimal threshold ($VIF < 0.1$), the arc will be colored green. The third layer of the glyph shows the MIC between each feature by the bar, and the heights of imply MIC values.

5.3. Feature Space and feature details

The Feature Space panel provides detailed information on features relation and their own characteristics. To represent various messages completely, we design the glyph which combines the pie chart and bar chart. The force-directed algorithm is applied to the layout of glyphs to relax the visual complexity. When the threshold of the Pearson Correlation Coefficient value changes, the link shows on between glyphs. Additionally, we utilize the Distance Correlation Coefficient value to fill the width of the line so that users can readily perceiving the connection between each feature.

While clicking the line between glyphs, as shown in Fig. 5, the Feature Detail panel shows their correlation and the detailed distribution. A density chart represented by a color gradient allows visualizing the combined distribution of two quantitative features. One feature is represented on the X-axis, the other one is on the Y-axis.

5.4. Data detail and feature subsets extraction

In the Data Detail panel, the system provides users with candidates for the audio dataset, showing the characteristics of the audio data in the form of histogram. On account of some features with small fluctuations, such as *zcr*, mapping them directly to the histogram is hard to extract effective information. We use density plots to show the distribution of characteristic numerical variables. To allow users to visualize the exact distribution of each feature, the interface supports users to view the distribution of each feature via a slider and assists users gets an preliminary perception.

Users utilize the Feature Selection and Feature Space panel to find the correlation between features and choose suitable feature subsets. They can click the checkbox to add or remove the feature interactively and click the button to extract the features. Users have access to click on one of the rows of the table to import the feature subset.

6. Case study

We use AFExplorer to conduct experiments and case studies to verify the effectiveness of our feature selection approach, and we also invite domain experts of machine learning and visualization to evaluate our system. In this section, we describe how AFExplorer can be used to explore the feature subset from 43 features for audio dataset.

In the beginning, since we start to use the system, the dataset view on the right provides us with a detailed distribution of the data. Multiple indicators in statistics which include standard deviation, maximum, and minimum, are presented at top of the bar chart. The Data Detail view can help us to obtain the effective information of the data in the least time. Here, we find that the *zcr* of the ID-2 feature has an extremely abnormal standard deviation (T3) as shown in Fig. 3. We decide to exclude this feature from the consideration of the feature subset. The best feature subsets of the three algorithms are stored in the table. We could directly click on the table to obtain the corresponding feature subset to explore in the Feature Selection panel.

In Feature Selection panel, based on heat map we found that the feature *zcr* is colored by pink in the STD line as shown in Fig. 3. It reveals that the three algorithms produce low importance value. Comparing with other features, such as energy, the color is also light (T1) (T3). In the heat map, we find that the inner-class features have the same appearance. In particular, in the feature of mfcc features, from mfcc-6 to mfcc-9 are adopted by LightGBM and XGBoost. The status is worthy for us to explore the hidden information.

In feature spaces, we choose features from mfcc-6 to mfcc-8 to compare the differences as shown in Fig. 5. Comparing with glyphs, the inside pie chart shows the importance of each algorithm. We observe the mfcc-8 has the maximum area. The second layer of the glyph represents its VIF value, it also has the smallest arc length. Refer to Fig. 4, mfcc-8 has infinitesimal collinearity. The third layer shows the MIC value of this feature and other features. In order to compare with other MIC values, the MI value with the feature itself is calculated and kept. The height of the histogram on mfcc-8's outer layer is almost invisible as shown in Fig. 6, indicating that the correlation with other features is low. Combining above three points, mfcc-8 has demonstrated excellent performance including high importance value, low correlation, and perfect stability.

We explore the similarity between the inner-class features by adjusting the threshold of the Pearson Correlation Coefficient. When the threshold is reduced to 0.3, mfcc-8 and mfcc-7 reveal a weak correlation(T2). By clicking on the line between the two glyphs, their detailed distribution and specific Pearson Correlation Coefficient values are displayed in the detailed view. In Fig. 5, these features IDs are saved in the table(T4). We use AFExplorer to select the feature subsets of three types of audio data from

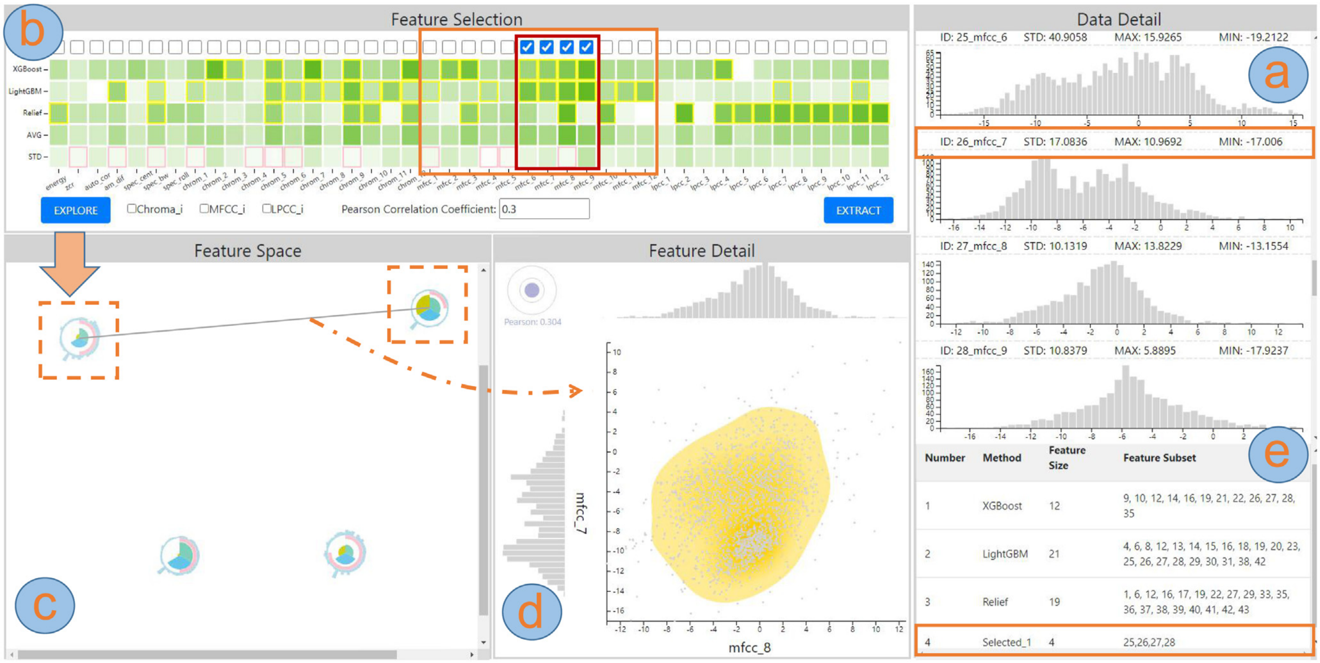

Fig. 5. Case of features extraction. (a) Data Detail gives detailed distribution about features. (b) Feature Selection presents the prominent features for users, we selected four features for exploration. (c) Feature Space displays the correlation with features. Through adjust the threshold of the Pearson Correlation Coefficient, the glyph will be connected by the gray line. (d) Feature Detail shows the mfcc-7 and mfcc-8 detail distribution. (e) Click the button of EXTRACT to export feature subsets.

Table 2
Results for the three types of datasets.

Dataset	Method	Feature subset	Dimension	Result		
				Accuracy	Precision	F1-score
Valve	XGBoost	1, 3, 4, 6, 7, 8, 11, 12, 13, 14, 15, 17, 18, 19, 20, 27, 28, 29, 31, 32, 33	21	**0.808**	0.875	0.813
	LightGBM	1, 4, 6, 7, 8, 11, 13, 14, 15, 16, 18, 20, 31, 32, 33, 34, 35, 40, 41	19	0.781	0.868	0.804
	Relief	1, 2, 3, 4, 5, 6, 15, 20, 23, 24, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42	27	0.712	0.78	**0.833**
	AFExplorer_1	1, 3, 11, 17, 18, 19, 20, 23, 24, 25, 26, 30, 36, 37, 39, 43	**16**	0.795	**0.892**	0.815
Fan	XGBoost	9, 10, 12, 14, 16, 19, 21, 22, 26, 27, 28, 35	**12**	0.738	0.776	0.846
	LightGBM	4, 6, 8, 12, 13, 14, 15, 16, 18, 19, 20, 23, 25, 26, 27, 28, 29, 30, 31, 38, 42	21	0.702	0.77	0.82
	Relief	1, 6, 12, 16, 17, 19, 22, 27, 29, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43	19	0.75	0.778	0.85
	AFExplorer_2	6, 12, 14, 16, 17, 19, 22, 25, 26, 27, 28, 29, 35, 38, 42	15	**0.768**	**0.801**	**0.861**
Toycar	XGBoost	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 26, 28, 30, 31, 33, 36, 37, 39, 40, 42	30	0.818	0.808	0.761
	LightGBM	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 23, 31, 32, 33, 36, 42	21	0.833	0.843	0.776
	Relief	2, 11, 13, 14, 16, 26, 28, 29, 30, 31, 32, 33, 35, 36, 38, 40, 41, 42	18	0.833	0.852	0.773
	AFExplorer_3	1, 5, 6, 7, 9, 22, 23, 25, 26, 27, 28, 35, 36, 38, 42	**16**	**0.862**	**0.864**	**0.821**

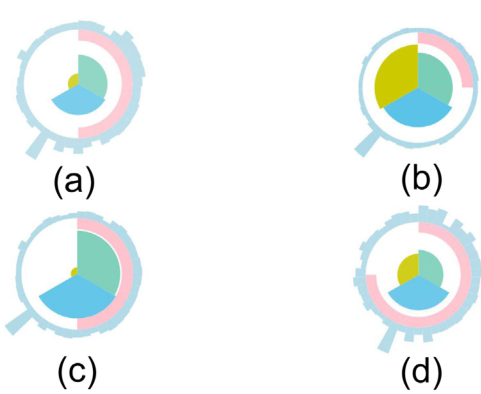

Fig. 6. (a) mfcc-7; (b) mfcc-8; (c) mfcc-9; (d) mfcc-6.

theToyADMOS and MIMII Dataset (Fan, Valve, and ToyCar). We apply the same model (AbaBoost) to test the selected feature subsets and applied accuracy, precision, and F1-score to these three commonly used indicators in machine learning for measuring test results. The results are shown in Table 2. The subsets of features selected by AFExplorer are smaller in number than the set of features retrieved by the three automated technologies and the accuracy and precision are enhanced.

7. Evaluation and discussion

We introduced our methods and visual analysis system of feature selection to three experts and asked for their opinions. The first expert (**E1**) is a researcher working in the machine learning field for four years. The second expert (**E2**) is a researcher from academic field that focuses on audio recognition. The third expert (**E3**) is a professor in computer science and a visualization researcher with experience in visualization and machine learning.

Usability and Interaction. We received much positive feedback and suggestive comments. **E1**, **E2**, and **E3** all agreed that our approach can relax the traditional feature selection on the problem of incomplete cognition on the features and assist them in efficiently extracting the feature subsets.

E3 affirmed the function of system perception and interaction. He said that from the perspective of user perception, the system successfully applied the contrast of color and transparency in the view to allow users to fetch more details. Meanwhile, new glyph has been designed to sustain users to further exploration. At the interactive level, users view the concrete details of the data and interact with the system through different controls, which helps to effectively explore and extract feature subsets. "Combining feature selection and visualization is a solid idea and makes the processing of extract feature be more impressive and effective. The visual analysis system can help users to conduct good feature sets and feature analysis in multiple views". **E2** said. He would like to employ such a system to conduct work related to different domains, such as music, animal sounds and soundscape ecology.

Scalability and Limitation. They also took notice of the system's efficiency. **E1** and **E2** stated the major concern that how to efficiently handle large datasets. Due to the characteristic of the REF algorithm, it will iteratively to choose the feature subsets step by step. Thus, our system has not yet been able to meet the real-time requirements. However, the system provides the view to assist users in exporting the feature subsets. The filter method and the wrapper method based on the decision tree adopted by the system have advantages in processing efficiency. When dealing with big data, the system faces the same problem in the field of machine learning, in that it takes time to train the data and generate the results. **E3** expressed the system should support more methods and algorithms for users to select, not limit to three algorithms that have been adopted. In the current work, we conducted only a few explorations and research on these three methods. In future work, we will consider integrating multiple methods to provide more choices. The system has limitations in dealing with unlabeled audio data and extracting features in a complicated environment. These directions are worthy of further exploration.

8. Conclusion

In this paper, we presented AFExplorer, a VA system with the aim to extract feature subsets using traditional feature selection methods interactive analysis and detection approaches. Various visualization panels support users in selecting applicable features and generating new feature subsets. Users can explore the impact of the features with several correlation coefficient measures and feature selection techniques. Interactive visualization techniques are proposed to display the features and help users visually compare the feature's differences with their own interests. Our evaluation demonstrated that the interface could be learned efficiently and the proposed workflow was comprehensible. Future developments should focus on fields with complex scenes and the combination of heterogeneous datasets from different sources. As future work, we intend to address aspects of scalability and support for greater diversity in data types. Taking advantage of the filter methods and visualization tools would be a nice choice in feature selection problems.

Ethical Approval

The study does not involve human subjects. All data used in the study are taken from public databases that were published in the past.

CRediT authorship contribution statement

Lei Wang: Conceptualization, Writing – original draft. **Guodao Sun:** Supervision, Writing – review & editing. **Yunchao Wang:** Data curation, Investigation, Visualization. **Ji Ma:** Guidance, Investigation. **Xiaomin Zhao:** Guidance, Proof reading. **Ronghua Liang:** Guidance, Reviewing, Proof reading.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is partly supported by National Key Research and Development Program of China (2020YFB1707700), National Natural Science Foundation of China (61972356, 62036009), and Fundamental Research Funds for the Provincial Universities of Zhejiang, China (RF-A2020001).

References

Agostini, Giulio, Longari, Maurizio, Pollastri, Emanuele, 2003. Musical instrument timbres classification with spectral features. EURASIP J. Adv. Signal Process. 2003 (1), 1–10.

Artur, Erasmo, Minghim, Rosane, 2019. A novel visual approach for enhanced attribute analysis and selection. Comput. Graph. 84, 160–172.

Battiti, R., 1994. Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Netw. 5 (4), 537–550.

Bogdanov, Dmitry, Wack, Nicolas, Gómez, Emilia, Gulati, Sankalp, Herrera, Perfecto, Mayor, Oscar, Roma, Gerard, Salamon, Justin, Zapata, José R., Serra, Xavier, 2013. Essentia: An audio analysis library for music information retrieval. In: International Society for Music Information Retrieval Conference (ISMIR'13). pp. 493–498.

Brossier, Paul M., 2006. The aubio library at mirex 2006.

Bullock, Jamie, 2007. Libxtract: A lightweight library for audio feature extraction. In: International Computer Music Conference Proceedings, Vol. 2007.

Chandrashekar, Girish, Sahin, Ferat, 2014. A survey on feature selection methods. Comput. Electr. Eng. 40 (1), 16–28.

Chen, Tianqi, Guestrin, Carlos, 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chiang, Leo H., Pell, Randy J., 2004. Genetic algorithms combined with discriminant analysis for key variable identification. J. Process Control 14 (2), 143–155.

Elmqvist, N., Dragicevic, P., Fekete, J.-D., 2008. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. IEEE Trans. Vis. Comput. Graphics 14 (6), 1141–1148.

Fang, Zheng, Guoliang, Zhang, Zhanjiang, Song, 2001. Comparison of different implementations of mfcc. J. Comput. Sci. Tech. 16 (6), 582–589.

Friendly, Michael, 2002. Corrgrams: Exploratory displays for correlation matrices. Amer. Statist. 56 (4), 316–324.

Gratzl, Samuel, Lex, Alexander, Gehlenborg, Nils, Pfister, Hanspeter, Streit, Marc, 2013. Lineup: Visual analysis of multi-attribute rankings. IEEE Trans. Vis. Comput. Graphics 19 (12), 2277–2286.

Guo, Diansheng, 2003. Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. Inf. Vis. 2 (4), 232–246.

Guyon, Isabelle, Elisseeff, André, 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Hamel, Philippe, Eck, Douglas, 2010. Learning features from music audio with deep belief networks. In: ISMIR. pp. 339–344.

Hancer, Emrah, Xue, Bing, Zhang, Mengjie, Karaboga, Dervis, Akay, Bahriye, 2018. Pareto front feature selection based on artificial bee colony optimization. Inform. Sci. 422, 462–479.

Janssens, Olivier, Slavkovikj, Viktor, Vervisch, Bram, Stockman, Kurt, Loccufier, Mia, Verstockt, Steven, de Walle, Rik Van, Hoecke, Sofie Van, 2016. Convolutional neural network based fault detection for rotating machinery. J. Sound Vib. 377 (377), 331–345.

Johansson, S., Johansson, J., 2009. Interactive dimensionality reduction through user-defined combinations of quality metrics. IEEE Trans. Vis. Comput. Graphics 15 (6), 993–1000.

Ke, Guolin, Meng, Qi, Finley, Thomas, Wang, Taifeng, Chen, Wei, Ma, Weidong, Ye, Qiwei, Liu, Tie-Yan, 2017. Lightgbm: a highly efficient gradient boosting decision tree. In: NIPS'17 Proceedings of the 31st International Conference on Neural Information Processing Systems, Vol. 30. pp. 3149–3157.

Khurana, Udayan, Samulowitz, Horst, Turaga, Deepak S., 2017. Feature engineering for predictive modeling using reinforcement learning.. In: AAAI. pp. 3407–3414.

Kira, Kenji, Rendell, Larry A., 1992. The feature selection problem: traditional methods and a new algorithm. In: AAAI'92 Proceedings of the Tenth National Conference on Artificial Intelligence. pp. 129–134.

Koizumi, Yuma, Saito, Shoichiro, Uematsu, Hisashi, Harada, Noboru, Imoto, Keisuke, 2019. 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.

Krause, Josua, Perer, Adam, Bertini, Enrico, 2014. Infuse: Interactive feature selection for predictive modeling of high dimensional data. IEEE Trans. Vis. Comput. Graphics 20 (12), 1614–1623.

Lartillot, Olivier, Toiviainen, Petri, 2007. Mir in matlab (ii): A toolbox for musical feature extraction from audio.. In: ISMIR. pp. 127–130.

Lewis, David D., 1992. Feature selection and feature extraction for text categorization. In: HLT '91 Proceedings of the Workshop on Speech and Natural Language. pp. 212–217.

Liu, Caifeng, Feng, Lin, Liu, Guochao, Wang, Huibing, Liu, Shenglan, 2021. Bottom-up broadcast neural network for music genre classification. Multimedia Tools Appl. 80 (5), 7313–7331.

MacEachren, A., Xiping, D., Hardisty, F., Guo, Diansheng, Lengerich, G., 2003. Exploring high-d spaces with multiform matrices and small multiples. In: IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714). pp. 31–38.

Michalak, Krzysztof, Kwaśnicka, Halina, 2006. Correlation-based feature selection strategy in classification problems. Int. J. Appl. Math. Comput. Sci. 16 (4), 503–511.

Mierswa, Ingo, Morik, Katharina, 2005. Automatic feature extraction for classifying audio data. Mach. Learn. 58 (2), 127–149.

Moffat, D., Ronan, D., Reiss, J.D., 2015. An evaluation of audio feature extraction toolboxes.

Muhlbacher, Thomas, Piringer, Harald, 2013. A partition-based framework for building and validating regression models. IEEE Trans. Vis. Comput. Graphics 19 (12), 1962–1971.

O'brien, Robert M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41 (5), 673–690.

Palaniappan, Rajkumar, Sundaraj, Kenneth, Ahamed, Nizam Uddin, 2013. Machine learning in lung sound analysis: A systematic review. Biocybern. Biomed. Eng. 33 (3), 129–135.

Peeters, Geoffroy, 2004. A large set of audio features for sound description (similarity and classification) in the cuidado project. CUIDADO IST Proj. Rep. 54, 1–25.

Purohit, Harsh, Tanabe, Ryo, Ichige, Kenji, Endo, Takashi, Nikaido, Yuki, Suefusa, Kaori, Kawaguchi, Yohei, 2019. Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. arXiv preprint arXiv:1909.09347.

Randall, Robert B., 2017. A history of cepstrum analysis and its application to mechanical problems. Mech. Syst. Signal Process. 97, 3–19.

Reshef, David N., Reshef, Yakir A., Finucane, Hilary K., Grossman, Sharon R., McVean, Gilean, Turnbaugh, Peter J., Lander, Eric S., Mitzenmacher, Michael, Sabeti, Pardis C., 2011. Detecting novel associations in large data sets. Science 334 (6062), 1518–1524.

Sánchez, Alberto, Soguero-Ruíz, Cristina, Mora-Jiménez, I., Rivas-Flores, Francisco Javier, Lehmann, D.J., Rubio-Sánchez, Manuel, 2018. Scaled radial axes for interactive visual feature selection: A case study for analyzing chronic conditions. Expert Syst. Appl. 100, 182–196.

Sharma, Garima, Umapathy, Kartikeyan, Krishnan, Sridhar, 2020. Trends in audio signal feature extraction methods. Appl. Acoust. 158, 107020.

Sindhwani, V., Rakshit, S., Deodhare, D., Erdogmus, D., Principe, J.C., Niyogi, P., 2004. Feature selection in MLPs and SVMs based on maximum output information. IEEE Trans. Neural Netw. 15 (4), 937–948.

Sousa, Ricardo, Antunes, Joel, Coutinho, Filipe, Silva, Emanuel, Santos, Joaquim, Ferreira, Hugo, 2019. Robust cepstral-based features for anomaly detection in ball bearings. Int. J. Adv. Manuf. Technol. 103 (5), 2377–2390.

Sun, Xin, Liu, Yanheng, Xu, Mantao, Chen, Huiling, Han, Jiawei, Wang, Kunhao, 2013. Feature selection using dynamic weights for classification. Knowl. Based Syst. 37, 541–549.

Sun, Guodao, Zhu, Sujia, Jiang, Qi, Xia, Wang, Liang, Ronghua, 2021. EvoSets: TRacking the sensitivity of dimensionality reduction results across subspaces. IEEE Trans. Big Data 1.

Turkay, C., Filzmoser, P., Hauser, H., 2011. Brushing dimensions - a dual visual analysis model for high-dimensional data. IEEE Trans. Vis. Comput. Graphics 17 (12), 2591–2599.

Wang, Yunhai, Li, Jingting, Nie, Feiping, Theisel, Holger, Gong, Minglun, Lehmann, Dirk J., 2017. Linear discriminative star coordinates for exploring class and cluster separation of high dimensional data. Comput. Graph. Forum 36 (3), 401–410.

Wang, Jinjiang, Ma, Yulin, Zhang, Laibin, Gao, Robert X., Wu, Dazhong, 2018. Deep learning for smart manufacturing: Methods and applications. J. Manuf. Syst. 48, 144–156.

Yang, Yi, Shen, Heng Tao, Ma, Zhigang, Huang, Zi, Zhou, Xiaofang, 2011. L2,1-norm regularized discriminative feature selection for unsupervised. In: Twenty-Second International Joint Conference on Artificial Intelligence.

Zeng, Zhihong, Pantic, Maja, Roisman, Glenn I., Huang, Thomas S., 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31 (1), 39–58.

Zhao, Jieqiong, Karimzadeh, Morteza, Masjedi, Ali, Wang, Taojun, Zhang, Xiwen, Crawford, Melba M., Ebert, David S., 2019. Featureexplorer: Interactive feature selection and exploration of regression models for hyperspectral images. In: 2019 IEEE Visualization Conference (VIS). pp. 161–165.

Zhou, Xinhui, Garcia-Romero, Daniel, Duraiswami, Ramani, Espy-Wilson, Carol, Shamma, Shihab, 2011. Linear versus mel frequency cepstral coefficients for speaker recognition. In: 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. pp. 559–564.