Article

A Feature-Reduction Scheme Based on a Two-Sample *t*-Test to Eliminate Useless Spectrogram Frequency Bands in Acoustic Event Detection Systems

Vahid Hajihashemi [1], Abdorreza Alavi Gharahbagh [1], Narges Hajaboutalebi [2], Mohsen Zahraei [3], José J. M. Machado [4] and João Manuel R. S. Tavares [4,*]

1 Faculdade de Engenharia, Universidade do Porto (FEUP), Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal; hajihashemi.vahid@ieee.org (V.H.); abalavi.gh@gmail.com (A.A.G.)
2 Department of Mathematics, Shahrood Branch, Faculty of Sciences, Islamic Azad University, Shahrood 3619943189, Iran; aboutalebi.n@yahoo.com
3 Department of Mathematics and Statistics, Faculty of Science, University of Regina, 3737 Wascana Pkwy, Regina, SK S4S 0A2, Canada; mzahraei326@gmail.com
4 Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal; jjmm@fe.up.pt
* Correspondence: tavares@fe.up.pt; Tel.: +351-22-041-3472

Abstract: Acoustic event detection (AED) systems, combined with video surveillance systems, can enhance urban security and safety by automatically detecting incidents, supporting the smart city concept. AED systems mostly use mel spectrograms as a well-known effective acoustic feature. The spectrogram is a combination of frequency bands. A big challenge is that some of the spectrogram bands may be similar in different events and be useless in AED. Removing useless bands reduces the input feature dimension and is highly desirable. This article proposes a mathematical feature analysis method to identify and eliminate ineffective spectrogram bands and improve AED systems' efficiency. The proposed approach uses a Student's *t*-test to compare frequency bands of the spectrogram from different acoustic events. The similarity between each frequency band among events is calculated using a two-sample *t*-test, allowing the identification of distinct and similar frequency bands. Removing these bands accelerates the training speed of the used classifier by reducing the number of features, and also enhances the system's accuracy and efficiency. Based on the obtained results, the proposed method reduces the spectrogram bands by 26.3%. The results showed an average difference of 7.77% in the Jaccard, 4.07% in the Dice, and 5.7% in the Hamming distance between selected bands using train and test datasets. These small values underscore the validity of the obtained results for the test dataset.

Keywords: acoustic event detection; Student's *t*-test; spectrogram; feature analysis; two-sample *t*-test

1. Introduction

In addition to the sense of sight, which is an important tool for interaction between humans and their environment, understanding the environment's acoustic signals is also crucial for survival. Using acoustic signals, one can often prejudge a scene or event without actually observing it. With the advancement of processors and AI systems, the proximity of machine learning (ML) algorithms to human perception, and the emergence of the smart city concept, urban smart monitoring systems based on audio and video have become increasingly popular. Considering the smart city concept, the automatic processing of audio and video data from urban areas enables city authorities to respond to incidents and quickly improve service quality. In congested urban areas, image-based event detection systems have long been used to monitor vehicle traffic or automatically detect events. These systems have the advantage of never getting tired, rarely making mistakes, and providing

comprehensive documentation of crimes and violations. Additionally, this topic is occasionally used in security systems. Under several circumstances, cameras cannot cover a scene; therefore, adding an acoustic signal can complement the monitoring system and enhance systems' accuracy and efficiency. So far, numerous methods have been proposed for processing and classifying urban events using the acoustic signal spectrogram, each with strengths and weaknesses. One aspect overlooked in spectrogram-based AED systems is the analysis of acoustic events based on similarities and differences of spectrogram bands related to different events separately. Considering this analysis, the trained system can use the spectrogram bands that differ between events to enhance its efficiency. Due to the lack of comprehensive research investigating the similarities and differences of spectrogram bands for different events, this article introduces a method to identify similar and dissimilar spectrogram bands for urban events from a mathematical and probabilistic perspective. In the meantime, the proposed method can be effectively used to select compelling features in the design and implementation of ML algorithms. The main advantages of the proposed method can be categorized into three areas:

- By analyzing the potential similarities and differences in the spectrogram bands of each urban event, one can focus on the similarities and dissimilarities within and outside the group of events, thus aiding in event classification.
- Since the proposed method uses a probabilistic model and confidence interval to evaluate the similarity and difference of spectrogram bands across different events, there is no need for heuristic methods. Also, mathematical analysis is more reliable than heuristic methods in ensuring accurate results.
- The proposed method can be used to identify irrelevant features in AED systems, and in cases where there is a high degree of similarity between two or more events, a secondary classifier can be designed using the outcomes of the proposed method to minimize errors.
- The proposed method can be used to identify useless bands in acoustic-based systems that use the spectrogram as a feature. Normality should be checked in each case to ensure the accuracy of the result.

The article is structured as follows: In Section 2, recent methods for AED and feature selection based on mathematical models are presented. Section 3 describes the proposed method. Section 4 is devoted to the results obtained by the proposed method and its comparison against other existing methods. Finally, a summary of the study is presented in the conclusion.

2. Literature Review

In recent years, multiresolution analyses, such as spectrograms, mel frequency cepstral coefficients (MFCCs), and wavelets, have been widely used in signal analysis and AED because of their suitability for finding patterns in time-varying signals. Hajihashemi et al. [1,2] used MFCC and wavelets for sound analysis in AED and acoustic scene classification. The authors also used wavelet scattering as another spectral feature in [1]. Roy et al. [3] used the spectrogram as a time-frequency expression of arterial Doppler signals to predict blood clots and microemboli. Several features were extracted from the spectrogram, such as the root mean of the local power spectrum and the modal frequency. Ibs-von Seht [4] aimed to provide an overview of volcanic activity using the spectrogram of seismic signals. Hafez et al. [5] predicted the timing of an earthquake using the spectrogram of signals obtained from the ground. Broussard and Givens [6] analyzed the oscillations of the posterior parietal cortex in rats and the impact of different acoustic signals on them using the spectrogram.

Liu et al. [7] employed the spectrogram in conjunction with the Hilbert–Huang Transform (HHT) for sleep apnea detection. Dennis et al. [8] used spectrogram and Hough transform features to detect acoustic events. Towsey et al. [9] used features extracted from the spectrogram to estimate the number of birds in a natural environment. Vales et al. [10] predicted earthquakes by analyzing data collected from the spectrogram of low-frequency

terms of terrestrial signals. Oliveira et al. [11] proposed an efficient method to detect bird activity using a spectrogram-based filter. The spectrogram separated the background sound from the bird's voice in this method.

Ghosh et al. [12] applied the spectrogram and Wigner–Ville transform of the vibration signal for vehicle detection. Xie et al. [13] introduced an AED system that used features extracted from spectrograms. Additionally, Xie et al. [14] used spectrograms, linear predictive coding, and MFCCs to estimate the number of frogs based on ambient sound. Using ambient sound, Sánchez-Gendriz and Padovese [15] analyzed biological choruses. In this study, features were extracted using the spectrogram, and an effective graphical expression for biological choruses was presented using the amplitude of the spectrogram in some frequency bands. Zhaoa et al. [16] separated the sounds of different bird species using MFCC, spectrogram, and an autoregressive model. Shervegar et al. [17] proposed a phonocardiogram spectrogram-based system for heart disease classification.

Nobre et al. [18] measured the biological parameters of caged domestic animals using an electric field. This research relied on the spectrogram to determine the frequency characteristics of long recordings. Ye et al. [19] used a combination of local and global features, including spectrogram entropy, to detect urban events. Goenka et al. [20] proposed a method for detecting seizures using quantitative electroencephalogram spectrograms. Hoyos-Barcelóa et al. [21] used local features of the acoustic signal spectrogram to detect coughs. The proposed method was implemented in a smartphone application and showed promising results. Waldman et al. [22] detected high-frequency oscillations within the human skull using electroencephalographic (EEG) signal spectroscopy. Yan et al. [23] used spectrograms to diagnose seizures based on a convolutional neural network (CNN) classifier whose input was spectral images.

Oliveira et al. [24] relied on the capabilities of spectrograms and ML methods, such as neural networks (NN) and support vector machines (SVM), to classify EEG signals and diagnose epilepsy. In addition to the spectrogram, cross-correlation and discrete Fourier transform were used in this study. Zhang et al. [25] employed acoustic sensors and a phase-sensitive optical time-domain reflectometer to distinguish five different acoustics. The authors extracted features using the spectrogram. Sahai et al. [26] considered spectrogram-related features for musical font separation. This method applied the spectrogram image as the input to the VGG network. Lin et al. [27] applied spectrogram features as the input to a deep neural network (DNN) and trained a semi-supervised CNN as an AED system in an urban area.

Spadini et al. [28] evaluated several acoustic features in detecting urban events, and the spectrogram was among them. Su et al. [29] used a two-stage CNN network to classify environmental acoustics considering features such as a log-mel spectrogram and MFCC-based features. Gloaguen et al. [30] proposed spectrogram features and non-negative matrix factorization to estimate road traffic levels. Satar et al. [31] proposed an AED method based on the spectrogram of data collected by the hydrophone. The continuous wavelet transform and spectrogram were used by Lapins et al. [32] to analyze seismic signals caused by volcanic activity. The audio spectrogram was among the acoustic features suggested by Vafeiadis et al. [33] for a smart home AED system. For environmental monitoring and counting of low detectable species, Znidersic et al. [34] used the spectrogram of an acoustic signal. Robinet et al. [35] used the spectrogram to extract transient noise characteristics in gravitational wave detectors.

Azab and Khasawneh [36] used the spectrogram to detect malware files. Kachaa et al. [37] analyzed the different conditions of dysarthric speech, which is a speech disorder related to muscle weakness, using the spectrogram of voice signals to interpret the different states of this disorder. Zeng et al. [38] extracted the spectrogram of arm movements and used this feature to classify the movements.

Franzoni et al. [39] proposed an emotion recognition system using a human voice spectrogram and a CNN-based classifier. Sinha et al. [40] extracted the audio spectrogram, and converted it to an image that was inputted into a CNN for audio classification.

Luz et al. [41] relied on different acoustic features, such as the spectrogram, to detect events in an urban space based on a CNN-based classifier. In analyzing the heart's electrocardiogram (ECG) signals, Gupta et al. [42] used the spectrogram. Manhertz and Bereczky [43] used the short-time Fourier transform (STFT) spectrogram to analyze vibration in a rotating electric machine and identify faults in early stages. In a study by Lara et al. [44], seismic and volcanic events were detected using a spectrogram and deep learning. Pham et al. [45] used a spectrogram-based method to classify scenes based on a CNN-DNN architecture.

Liu et al. [46] combined convolutional recurrent neural networks and mel spectrogram, delta, and delta-delta features for underwater target recognition. Kadyan and Bawa [47] proposed a two-level augmentation scheme via the spectrogram of speech signals using transfer learning techniques for an automatic speech recognition system. Pahuja and Avijeet [48] proposed a bird sound-based recognition system for classifying eight species of birds using the STFT spectrogram for feature extraction. Zhang et al. [49] used gradient-weighted class activation mapping as a CNN and the mel spectrogram as a feature for acoustic scene classification. Cheng et al. [50] suggested a spectrogram-based sound recognition system using AlexNet, which identified passing vehicles with modified loud exhausts. Wang et al. [51] used the spectrogram of underwater signals for time-frequency tracking and enhancement of whistle signals. The application of whistle signals is in research about cetaceans. You et al. [52] used audio spectrogram transformers and a CNN to generate embeddings for the few-shot learning of bioacoustic AED. Bhangale et al. [53] combined mel spectrogram and other acoustic features and used them as input for a parallel emotion network for speech emotion recognition. Özseven [54] discussed the effectiveness of the spectrogram as a time-frequency domain image in urban sound classification. Latif et al. [55], Shafik et al. [56] and Mushtaq et al. [57] used the spectrogram as an effective acoustic feature in speech emotion recognition, speaker identification, and environmental sound classification, respectively. All of these approaches were based on deep learning.

The spectrogram has also been used in many medical applications, such as cough detection [58], detection of cardiovascular disease and epilepsy using ECG and EEG signals [24,59], sleep spindles [60], and scalp peak ripples [61] detection using EEG signals. It has also been used for sleep apnea–hypopnea syndrome diagnosis based on nasal airflow signal [62] and snoring detection system using voice [63]. In industrial applications, the spectrogram has been used for fault detection using vibration signals [64], fault detection in gearboxes based on sound [65], and fault detection in rotary systems using data from various sensors [66]. Verification of bird diversity [67] and the detection of shoots in forests [68] using ambient sound, seismo-acoustic event prediction using vibration signals and ground waves [69], and an AED system [70] are other state-of-the-art applications that use the spectrogram as a feature extraction method.

In most of the studies mentioned above, the classifier employed was a DL network. According to the reviewed studies, the following findings were observed:

- In some methods, the spectrogram image was used as the input to a two-dimensional (2D) DNN;
- Various features extracted from the spectrogram are used as the input;
- In some cases, experts have analyzed spectrogram images to distinguish between different conditions;
- Based on our best search, no quantitative methods have been proposed to determine which frequency bands of the spectrogram are most effective for the application under study.

In the current study, an efficient method to separate the useful from useless bands of the spectrogram regardless of the used classifier was developed based on statistical tests. The proposed method can be used to determine the similarities and differences in the frequency bands of the spectrogram considering different classes. Using the results of the proposed method, the noise is reduced by removing similar frequency bands from the spectrogram, and the accuracy and learning speed are increased. Therefore, in an AED system, the proposed method can provide insights into the different events associated with

the spectrogram frequency bands relative to the background and increase the system's accuracy and speed.

3. Materials and Methods

This section provides an overview of the dataset used in this study and explains the architecture of the proposed system. In addition, it gives the theoretical background for each step of the proposed method.

3.1. Dataset

As one of the most popular datasets used in AED studies, the well-known public URBAN-SED dataset was also used in the current study. This dataset contains ten sound events, as shown in Figure 1: air conditioner, car horn, children playing, dog bark, drilling, engine idling, gunshot, jackhammer, siren, and street music. In addition, there is an 11th class, defined as the background, which does not include any of the mentioned events.
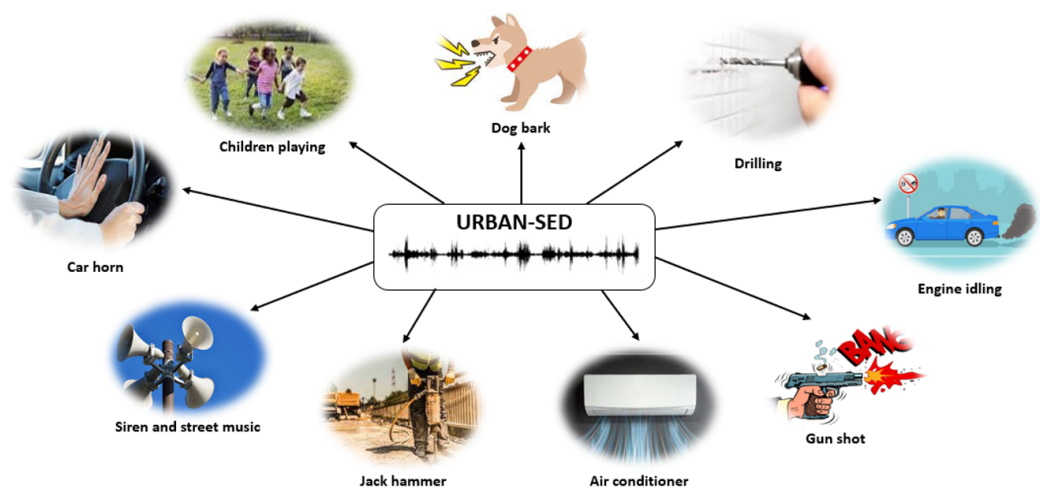
Figure 1. URBAN-SED dataset: included sound events.

The URBAN-SED dataset was generated using the Scaper library for synthesizing and augmenting acoustic scenes and developed by incorporating background noise into the original sounds sourced from the UrbanAcoustic8K dataset. The UrbanAcoustic8K dataset contains 8732 trimmed acoustic clips recorded in an urban environment, with the most extended clip lasting four seconds. To generate the URBAN-SED dataset, background noise was added to the original sounds, and the duration of each sample was set to ten seconds. Since the UrbanAcoustic8K sounds were recorded in a natural environment, several events can co-occur. To standardize the comparison between the different studies tested on the URBAN-SED, the data were divided into three categories: training, testing, and validation. The training set consists of 6000 samples, whereas the remaining two contain 2000 samples each.

3.2. Analysis of Spectrograms

The current study aims to separate the useful bands from the useless ones of the spectrogram in a spectrogram-based classification system. The spectrograms of acoustic events are analyzed to identify frequency bands that vary among different events to achieve this objective. These bands can be used as indicators of acoustic events. Figure 2 provides an overview of the proposed methodology, and its pseudocode is outlined in Algorithm 1.

Assume an audio sample is divided into N frames with equal length, and X_M is the spectrogram of each frame, so $X_{M \times N}$ is the spectrogram matrix of the sample. Each column of X denotes a frame number, i.e., a time interval, while each row represents a frequency band. Suppose that the only difference between different time intervals, i.e., columns of the spectrogram matrix, is the presence of an acoustic event in that column. In this

case, the columns can be divided into two non-overlapping groups: one with and one without an acoustic event. To avoid ambiguity, overlapping periods with more than one event were removed from the input data (Figure 3). In statistics, the two-sample *t*-test checks the equality of means between two populations, i.e., two groups, that follow a normal distribution.

In this study, each frequency band was assumed to be a population. The normality condition was checked and validated in each frequency band to ensure the validity of the final result. Thus, if the equality of means between the two populations is rejected, it is reasonable to mark the frequency band as usable in the AED. In contrast, if the test for equality of means is accepted, frequency bands are useless due to similarity. To mitigate the impact of transient and short-term noise on the results, all training samples were used for analysis, and the results were averaged. Performing this analysis for all acoustic events in the dataset and then averaging the results yields a comprehensive model of the similarities and differences between acoustic events across the entire frequency band.
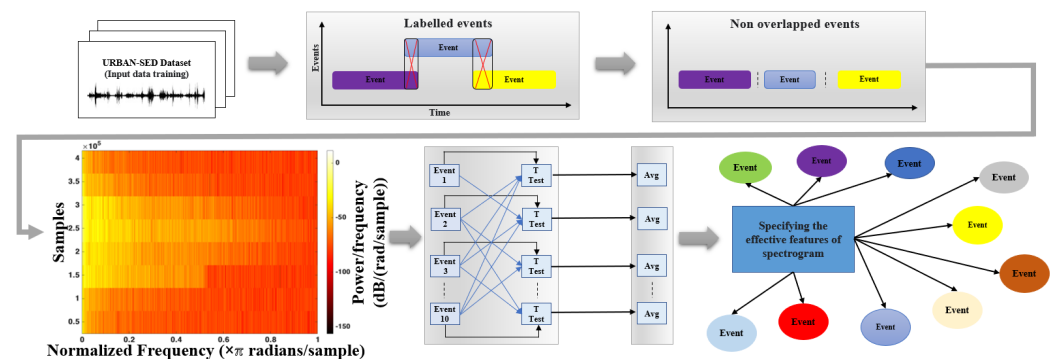

Figure 2. Overview of the proposed method.

Algorithm 1 Pseudocode of the proposed method

Require: A dataset including audio and time tags of events
 1: **for all** acoustic signal in the dataset **do**
 2:     Extract the audio spectrogram $X_{M \times N}$, where M is the number of frequency bands and N denotes the time interval
 3:     Divide the spectrogram into two groups: X_e—Signal with an acoustic event, and X_{Ne}—Signal without an event, such that X_e is $M \times U1$ and X_{Ne} is $M \times U2$ where $U1 + U2 = N$
 4:     **for all** row in the spectrogram (Frequency bands 1 to M) **do**
 5:         Each row of X_e and X_{Ne} is assumed as a population with the number of $U1$ and $U2$ samples and called S_e and S_{Ne}, respectively
 6:         Perform a data normality test on S_e and S_{Ne} and store the result
 7:         Perform a mean equality test on S_e and S_{Ne} and store the result
 8:     **end for**
 9: **end for**
10: **Output 1:** Accept or reject data normality and mean equality test in all frequency bands for each input.
11: **Output 2:** Percentage of acceptance/rejection of the data normality and mean equality test in all input frequency bands.
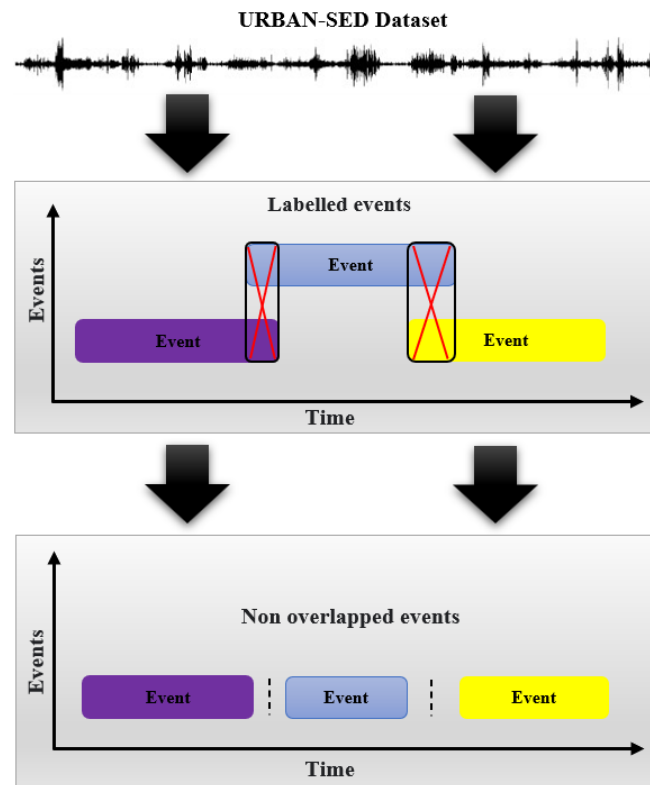

Figure 3. Process of removing time intervals containing multiple events and preparing populations.

3.2.1. Mel Spectrogram

Many studies on audio analysis and AED used the mel spectrogram and MFCC as features. The current study analyzed mel spectrograms to identify the frequency bands that are effective for AED systems. Knowing the effective frequency bands for each acoustic event in the spectrogram makes it possible to annotate the presence and even the start and end times of the acoustic event with greater accuracy than considering all frequency bands. Moreover, identifying the dominant frequency bands for each event allows for separating similar but not identical events, thereby reducing the possibility of errors. Several parameters need to be set in the mel spectrogram, including the number of filters in the filter bank, the duration of the time interval, the type of window used in the time domain, the number of points in the Fourier transform, and the amount of overlap between time intervals. These parameters affect the system's accuracy. Figure 4 shows a mel spectrogram filter bank.

It is well known that the bandwidth of mel filters is smaller at low frequencies and increases at higher frequencies. In the analysis performed in this study, an audio sampling frequency of 44,100 Hz was used, and the number of points in the time window was set to 2048, with an overlap of 1024 points. This means that each time window overlaps adjacent time slots by half of the window size. Also, the number of mel filters was set to 173. Different windows in the time domain have two important but opposite characteristics: a narrow "main lobe width" and a high attenuation of the side lobes. Compared to Hanning and Hamming, the Blackman–Harris window has a wider main lobe width, which is a disadvantage but without effect in the current application; however, it has a stronger sidelobe attenuation than other windows, which is highly desirable, so Blackman–Harris was used in this study [71]. Based on the fact that all clips in the dataset have a duration of ten seconds, the final mel spectrogram matrix is a 173×429 matrix, where 173 frequency bands correspond to rows and 429 columns correspond to time intervals. In this step, based on the dataset labeling, intervals where only one audio event is present can be separated from intervals without events. To avoid ambiguity, the intervals where the acoustic event started or ended were excluded from the analysis.
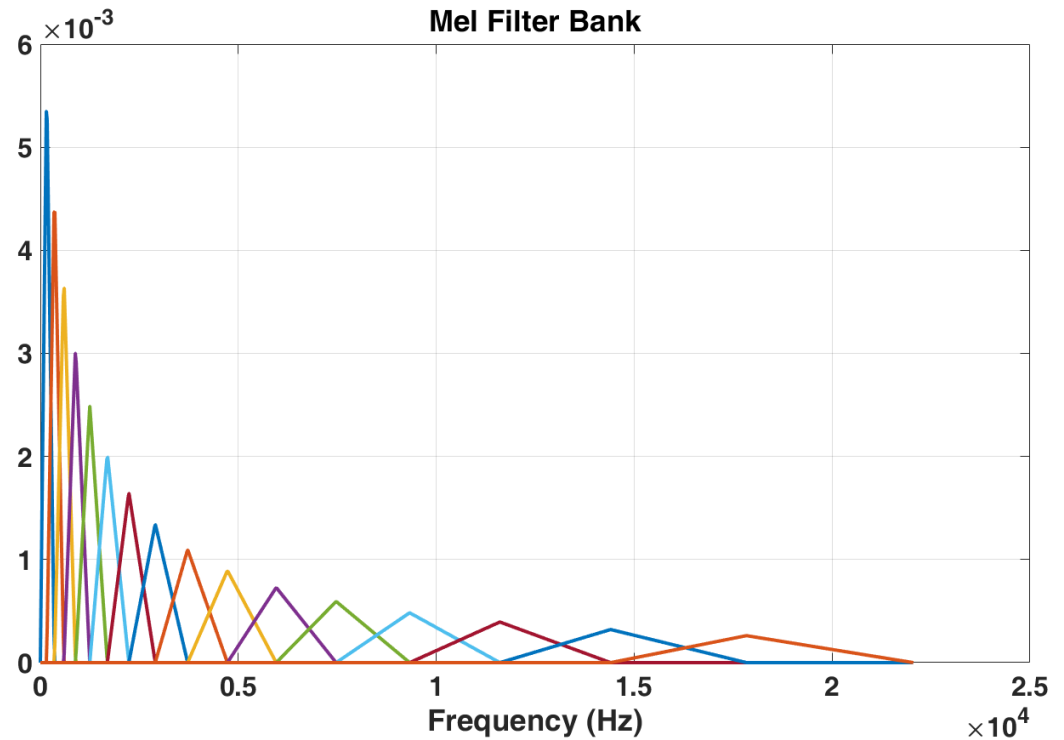
Figure 4. Example of a mel spectrogram filter bank (each line represents one of the filters used).

3.2.2. Student's *t*-Test Analysis

To verify the effectiveness of each frequency band in detecting audio events, a two-sample Student's *t*-test is performed, as previously mentioned. It is assumed that the spectrogram values within the frequency bands at different time intervals follow a normal distribution (details of the normality test are discussed in the following section). Frequency bands can be divided into two parts, with/without events, considering their time intervals, and the means of these two parts, as two populations, can be compared using the Student's two-sample *t*-test, which is dependent on the variance of populations. If the variances of the two populations are known, the test statistic is given by [72]:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{1}$$

where σ_1, σ_2 are standard deviations, and n_1, n_2 are the number of samples, i.e., the number of time intervals, in each population. In cases where the variances are unknown but assumed to be equal, the test statistic is as [72]:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{2}$$

where S_p^2 is the pooled variance, which can be calculated as [72]:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \tag{3}$$

where S_1^2 and S_2^2 are the variances calculated from two populations and n_1, n_2 are the number of samples as in Equation (1). If the variances are unequal, the following equation [73] can be used:

$$t' = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \tag{4}$$

where S_1^2 and S_2^2 are the variances of the two populations, and n_1, n_2 are the number of samples. It was verified that all populations belonging to all frequency bands have equal variances. The results indicated that, in most cases, the variances were not equal. Therefore, Equation (4) was applied to perform the *t*-tests. Finally, supposing that the mean equality test of two populations, one with an event and another without an event, in a frequency band, is accepted, this frequency band is ineffective for detecting this event because the values of this frequency band remained the same with and without the event. In contrast, rejecting the test shows that this frequency band differs with and without events, which makes it helpful in detecting the event. In cases where the test is accepted for some clips and rejected for others, this frequency band can be used to indicate the event; however, it is not a strong indicator. This statistic is highly efficient in feature selection because it can determine the usefulness of a feature in a binary or multi-event classification system, regardless of the classification method. To the best of our knowledge, there have been no previous studies on the functionality of the frequency bands of the spectrogram in AED systems, and the current study was the first attempt to explore the effects of each spectrogram frequency band on AED systems.

3.2.3. Normality Tests

The normality of two populations is a necessary assumption in a test of the equality of two means. Various tests can be used to assess data normality. Five tests were used in this study: Kolmogorov–Smirnov, Lilliefors, Anderson–Darling, Jarque–Bera, and Shapiro–Wilk. The Kolmogorov–Smirnov test is a nonparametric test that examines the fit of a given probability distribution to a set of samples. This test first transforms the data into the standard normal form: zero mean, unit variance. Subsequently, the cumulative distribution function of the data is compared to a standard normal cumulative distribution function. The normality of the data can be accepted or rejected based on the differences between the two graphs. With some modifications, this test is also used to check the goodness of fit. The Lilliefors test is similar to the Kolmogorov–Smirnov test in its initial stage. The difference between the two tests is how the cumulative distribution function is calculated. In the Lilliefors test, the data is not transformed into standard form, and the cumulative distribution is calculated directly. Normality is accepted or rejected based on the maximum discrepancy between the ideal normal cumulative distribution and the empirical cumulative distribution function of the data. One challenge of this test is determining the significance of the difference between the data distribution function and the ideal form. Because the test function is calculated based on the mean and variance of the data, it appears to be similar to the normal function, which can be considered a weakness. Nevertheless, this test can yield better results in some cases than the Kolmogorov–Smirnov test. The third test is the Anderson–Darling. In the general form, the Anderson–Darling test compares any population to any possible distribution, including the normal distribution. Similar to the Kolmogorov–Smirnov test, this test involves comparing the empirical distribution function of the data to the ideal normal distribution function. However, the initial assumptions of the Anderson–Darling test differ. The Anderson–Darling test has four different modes for testing the normality of data, which are as follows:

1. The mean and variance of the data are both known;
2. The data variance is known, but the mean is unknown;
3. The mean of the data is known, but the variance is unknown;
4. Both the mean and variance of the data are unknown.

In the current study, the mode where the data's mean and variance were unknown was used. In such cases, the mean and variance of the data are first estimated using

statistical relationships. The data are then transformed into a standard form according to the following relationship [73]:

$$z = \frac{x - \mu}{\sigma}.$$ (5)

The following equation was used to estimate the cumulative distribution function of the standardized data [73]:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)(\ln \phi(Y_i) + \ln(1 - \phi(Y_{n+1-i}))).$$ (6)

Based on this statistic, the following statistic is estimated [73]:

$$A^{*2} = A^2 \left(1 + \frac{4}{n} - \frac{25}{n^2} \right).$$ (7)

It is important to note that this relationship is valid when the mean and variance are unknown and are estimated based on the data. If any of the A^2 or A^{*2} values exceed the value given in the Anderson–Darling distribution table, the assumption of data normality is rejected. The fourth test performed to ensure data normality is the Jarque–Bera test. Unlike previous tests, this test compares the data probability distribution with a standard normal distribution based on skewness and kurtosis. Deviations of skewness and kurtosis from the normal distribution values lead to the rejection of normality. If the mean and variance of the data are not known, skewness and kurtosis can be calculated using the following equations [74]:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{3/2}},$$ (8)

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^2}.$$ (9)

After calculating skewness and kurtosis as the third and fourth central moments of the data, the Jarque–Bera statistic is calculated as [75]:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4}(k - 3)^2 \right),$$ (10)

where n is the number of samples. To accept or reject normality, the Jarque–Bera statistic is compared with the Jarque–Bera table obtained by the Monte Carlo method or chi-square approximation. Here, the Monte Carlo table was used based on the number of samples in the two populations. According to [76,77], the Shapiro–Wilk test is the most appropriate normality test for data with a sample size of less than 50:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_I \right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$ (11)

where x is the samples, $\bar{x}$ is the mean, and the (a_i) coefficients are normalized best linear unbiased estimators that can be computed using methods such as the Monte Carlo method [78,79]. Because of the considerable variation in population size and the dependence of the normality test accuracy on the number of samples, the types of tests in this study were selected based on the number of samples [80–82].

3.3. Validation Scheme

Figure 5 illustrates the scheme used to validate the results. In the first step, effective bands were determined between events using the train and test data separately. Then,

the values 1 (one) and 0 (zerp), were assigned to effective and excluded bands, respectively, and a binarized vector with 173 elements for each pair of events was created.
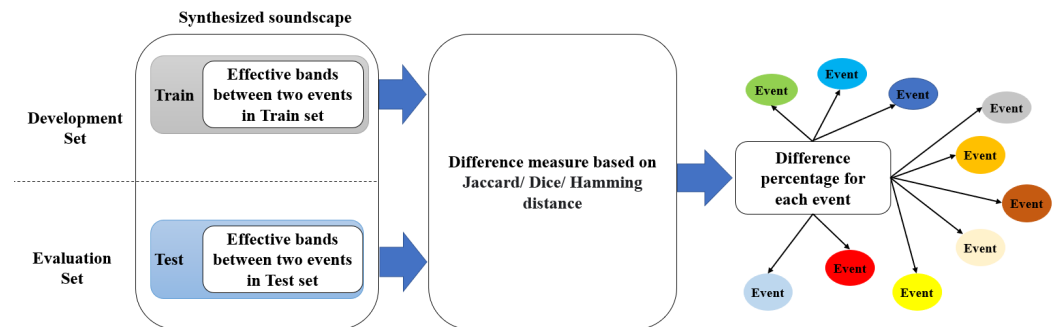
Figure 5. Scheme used for the validation of the results.

In the second step, the Dice coefficient, Hamming distance, and Jaccard distance were used. Among these metrics, the Dice coefficient measures similarity, while the Hamming and Jaccard distances measure differences. In the proposed scheme, the 1—Dice coefficient was used as the Dice distance to measure differences. In an ideal scenario, the results of the training and testing data are perfectly similar, so the Dice, Hamming, and Jaccard distances should all be zero. Given two binarized vectors, R_{train} and R_{test}, each with n binary elements, the Jaccard distance measures the missed overlap between R_{train} and R_{test} relative to the total number of bands, regardless of the excluded bands. First, the following parameters were defined:

- E11—number of elements where both R_{train} and R_{test} are equal to 1 (one);
- E01—number of elements where R_{train} is equal to 0 (zero) and R_{test} to 1 (one);
- E10—number of elements where R_{train} is equal to 1 (one) and R_{test} to 0 (zero);
- E00—number of elements where both R_{train} and R_{test} are equal to 0 (zero).

Each binary element must fall into one of these four parameters, meaning that:

$$E11 + E00 + E10 + E01 = \textit{Total spectrogram bands}, \tag{12}$$

where *Total spectrogram bands* is equal to 173 in the current study. On one hand, the Jaccard distance, d_J, is given by [83,84]:

$$d_J = \frac{E01 + E10}{E01 + E10 + E11} \tag{13}$$

The Hamming distance measures the missed overlap between R_{train} and R_{test} relative to the total number of bands and is given by [84,85]:

$$d_H = \frac{E01 + E10}{E01 + E10 + E11 + E00}. \tag{14}$$

On the other hand, the Dice distance is defined as [83,84,86]:

$$d_D = \frac{E01 + E10}{E01 + E10 + 2E11}. \tag{15}$$

Since the Dice distance does not satisfy the triangle inequality, it can be considered a semi-metric version of the Jaccard distance. All metrics are reported here as percentages.

4. Results and Analysis

In this study, when the number of populations was less than 50, the Shapiro–Wilk test was used for the normality test. When samples exceed 50, alternative tests are recommended to verify normality [87]. In these cases, the dominant response of the Liliefors,

Anderson–Darling, and Jarque–Bera tests, were used. Almost all the statistical tests had a reasonable response when the number of samples exceeded 300. In this case, the dominant response of the four tests, Lilliefors, Anderson–Darling, Jarque–Bera, and Kolmogorov–Smirnov, was chosen as the result, which is the typical case here. Table 1 presents the normality test results of the four normality tests in the most common issue in the current study. The selected normality test results are given in the hybrid column of Table 1. The results in Table 1 confirm the validity of the assumption of normality.

Table 1. Data normality test results (KS—Kolmogorov–Smirnov test, LF—Lilliefors test, AD—Anderson–Darling test, and JB—Jarque–Bera test).

| Event Category	KS	LF	AD	JB	Hybrid
Air conditioner	96.70	80.53	77.92	88.84	85.33
Car horn	99.40	95.43	77.70	95.93	93.07
Children playing	87.21	87.70	82.83	76.59	82.47
Dog bark	98.06	85.42	74.86	75.89	83.38
Drilling	99.07	93.10	76.32	80.91	99.70
Engine idling	89.36	87.36	81.06	82.51	89.36
Gunshot	93.52	80.86	67.82	74.38	82.42
Jackhammer	99.70	84.48	78.61	85.91	98.53
Siren	92.85	91.84	77.38	87.95	94.61
Street music	98.99	88.01	76.49	82.16	96.92

4.1. Mean Equality Test

To perform the two-sample test of means, the following assumptions were considered:

- The frequency bands of the spectrogram examined in this study were 173; the mean equality test was performed separately for each frequency band;
- There was only one event in the populations selected for the test;
- The minimum number of samples in each population was equal to nine;
- The populations had an unequal number of samples;
- Each population, which consisted of consecutive samples belonging to an event, was compared with events from the same audio file to minimize the effects of background noise;
- The percentage of rejections in the "mean equality test" was calculated separately for each audio event compared to other events and background, i.e., no event, using all training samples (6000 samples);
- The assumed confidence interval for all tests was equal to 95%;
- If a population failed in the normality test and its skewness and kurtosis deviated strongly from the normal distribution, it was excluded from the test;
- The higher values in Figures 6–8 indicated frequency bands with a higher probability of a mathematical difference between two acoustic events, as indicated by a higher percentage of rejections in the mean equality test.

In Figure 6, four events are depicted according to the background: gunshot, jackhammer, siren, and street music. It can be seen that the jackhammer spectrogram differs from the background in the bands between 10 and 150 in at least 80% of the clips. However, this situation is not observed for the other three events. Among the four events, siren differs from the background only in a relatively narrow range of frequency bands. Figure 7 depicts the test result of the mean equality test between the gunshot event and other events. It is possible to perceive that the importance of different frequency bands in distinguishing the gunshot event from other events varies depending on the type of the second event. In all the reported results, frequency bands with a higher rejection percentage of the mean

equality test (value of the vertical axis of the graphs depict in the figures) are more suitable for classification.
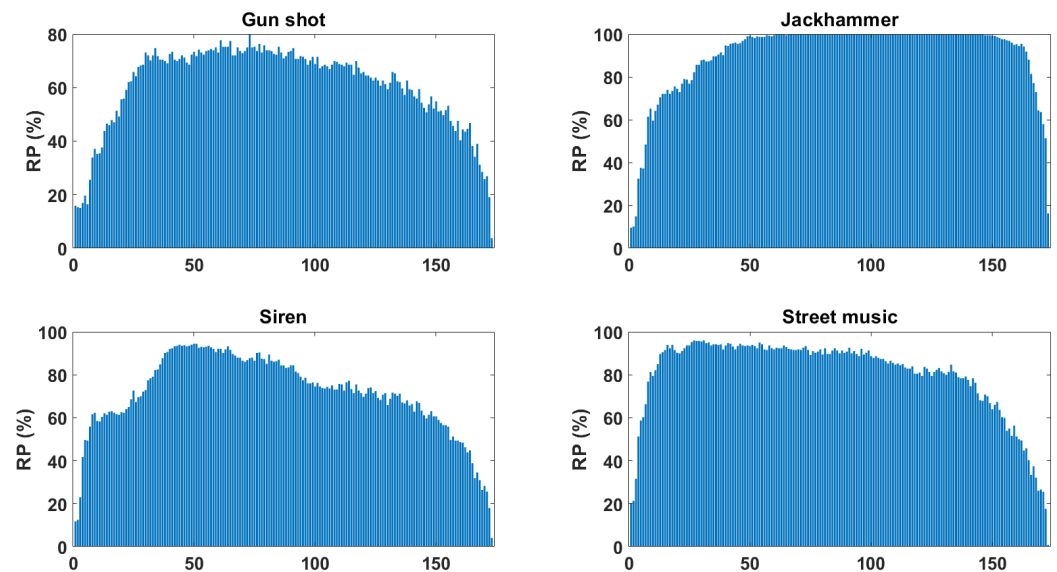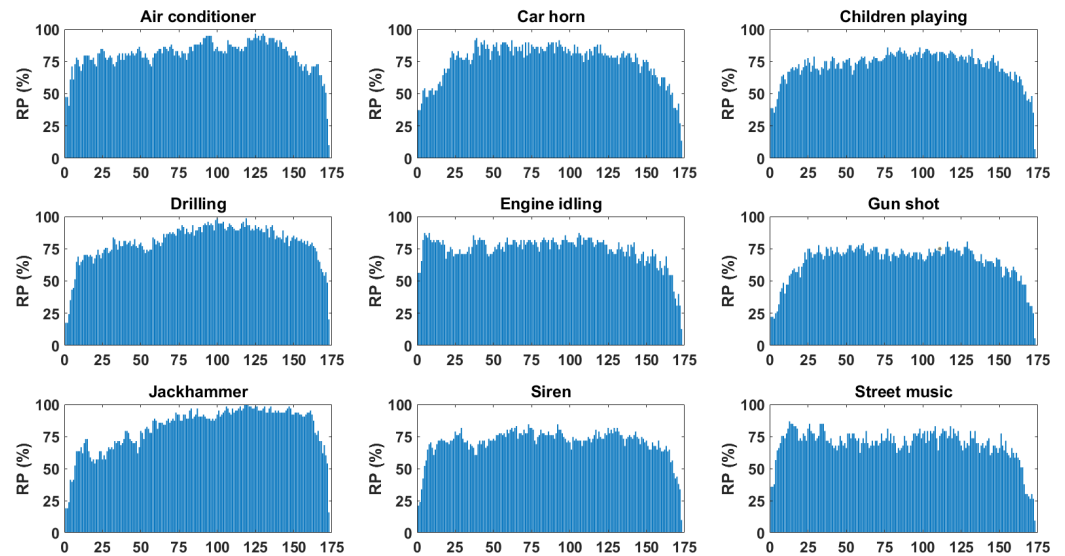

Figure 6. Rejection percentage (RP) of mean equality test for four events according to the background.

Based on the results depicted in Figure 7, among the events, dog bark, siren, and street music have a smaller area under the curve than the others with gunshot, which indicates a higher probability of classification error between these three events and gunshot when classified by the spectrogram. Conversely, the air condition, engine idling, and jackhammer exhibited the most significant differences. Thus, if an AED system's confusion matrix shows significant errors between the gunshot and dog bark classes, a new classifier can be developed using the most appropriate bands from the spectrogram, as shown in Figure 7. This approach enhances the AED efficiency and reduces errors. Similar analyses can be performed for other events. For example, Figure 8 shows the results for dog bark according to the other events, which differs from the gunshots. As an implicit rule, when the rejection percentage of the mean equality test is less than 75%, the frequency bands are considered ineffective.
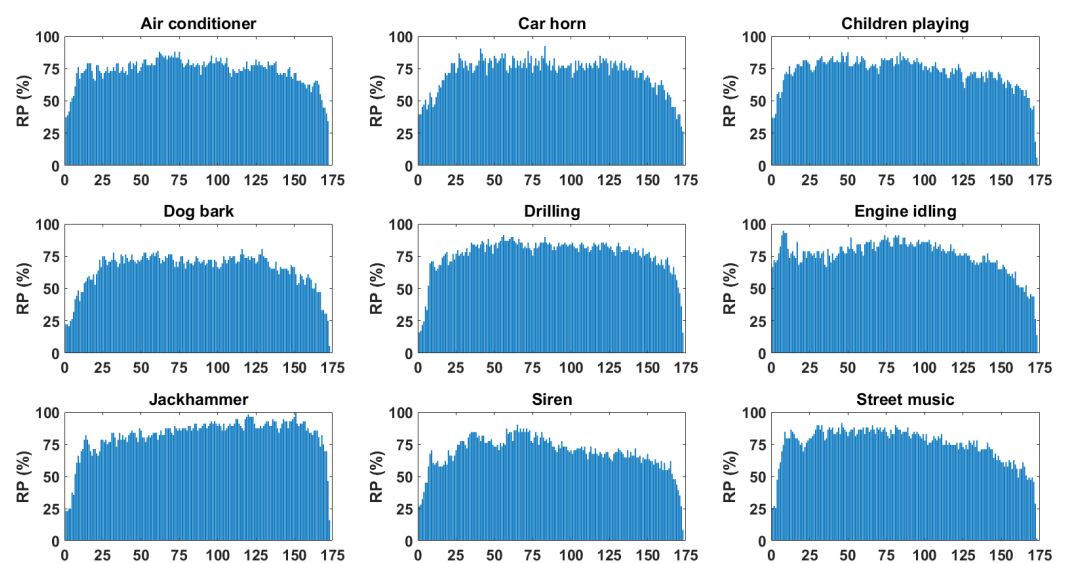
Figure 7. Rejection percentage (RP) for mean equality test between gunshot and other events.

Figure 8. Rejection percentage (RP) for mean equality test between dog bark and other events.

This statistical criterion can be used as an efficient method for feature selection based on statistical patterns without the need for evolutionary or iterative techniques. The only limitation of the proposed method is the requirement for many samples. The ratios between the effective frequency bands, i.e., with a rejection percentage greater than 75%, and the total frequency bands are indicated separately in Table 2. A higher ratio indicates that more spectrogram bans can be helpful. In contrast, a smaller ratio indicates that more spectrogram bans can be removed in developing an AED system. The weakest result in Table 2 is 17.9% (between siren and dog bark), indicating that only 31 of the 173 spectrogram frequency bands effectively distinguish between the dog bark and siren classes. Regarding the weak features, the dog bark has more in common with other events, showing that for this event, many spectrogram bans can be removed during the classifier design without reducing the efficiency. According to the results in Table 2, many spectrogram bands (approximately 26.3%) can be omitted during the AED design. Thus, in addition to reducing noise, complexity, and training time, the number of samples required to train the system is reduced.

Table 2. Percentage of effective bands to total spectrogram bands in the train set of the URBAN-SED dataset (the listed events are: A = air conditioner, B = car horn, C = children playing, D = dog bark, E = drilling, F = engine idling, G = gun shot, H = jackhammer, I = siren, J = street music).

Class	A	B	C	D	E	F	G	H	I	J
A	-	87.86	78.61	78.03	93.06	84.39	80.35	90.17	91.91	84.97
B	87.86	-	71.10	58.38	80.92	80.92	48.55	83.24	62.43	75.14
C	78.61	71.10	-	65.90	84.97	88.44	69.94	86.71	79.19	71.68
D	78.03	58.38	65.90	-	70.52	87.28	32.95	86.71	17.34	53.18
E	93.06	80.92	84.97	70.52	-	93.06	73.41	91.33	78.03	87.86
F	84.39	80.92	88.44	87.28	93.06	-	82.08	91.33	82.08	84.97
G	80.35	48.55	69.94	32.95	73.41	82.08	-	82.08	45.09	45.09
H	90.17	83.24	86.71	86.71	91.33	91.33	82.08	-	87.28	80.35
I	91.91	62.43	79.19	17.34	78.03	82.08	45.09	87.28	-	76.30
J	84.97	75.14	71.68	53.18	87.86	84.97	45.09	80.35	76.30	–
Background	87.9	67.1	74.0	23.1	76.9	83.2	36.4	87.9	39.9	72.3

4.2. Validation

Figure 5 illustrates the scheme used to validate the results in Table 2. If the results of Table 2 are valid, the selected bands from the test samples would be relatively similar to the training samples.

Based on the results of Table 3, it can be seen that in the Jaccard metric the average difference in all events is 7.77%, and the greatest change occurred for the dog bark and siren events. This great difference (25%) is due to the low number of effective bands between the dog bark and siren events (Table 2) as E_{11} and its effect on the denominator of Equation (13) does not necessarily indicate a high mismatch. The Dice and Jaccard metrics, due to the denominator of Equations (13) and (15), when the number of effective bands between two events E_{11} is small, may also show a high value in the low mismatch. According to the results of Table 2, only 17.34% of bands (equivalent to 30 bands) between these two events were effective. In this situation, a slight mismatch of ten bands (out of 173 bands) between selected bands in training and testing data, showed a 25% mismatch in the Jaccard metric. In such cases, bands with values close to but less than the specified rejection percentage of the mean equality test can be selected as effective bands.

Table 3. Jaccard distance (%) between the results of train and test datasets (the listed events are: A = air conditioner, B = car horn, C = children playing, D = dog bark, E = drilling, F = engine idling, G = gun shot, H = jackhammer, I = siren, J = street music).

Class	A	B	C	D	E	F	G	H	I	J
A	-	5.13	6.43	5.88	4.91	5.33	7.69	8.13	5.56	6.67
B	5.13	-	7.52	7.41	5.41	8.11	11.58	6.04	8.47	5.22
C	6.43	7.52	-	6.56	8.92	7.69	6.20	5.23	7.59	6.20
D	5.88	7.41	6.56	-	8.96	5.81	13.64	6.49	25.00	11.00
E	4.91	5.41	8.92	8.96	-	6.75	9.93	5.03	7.59	5.77
F	5.33	8.11	7.69	5.81	6.75	-	6.85	4.97	9.52	7.19
G	7.69	11.58	6.20	13.64	9.93	6.85	-	7.59	11.36	13.48
H	8.13	6.04	5.23	6.49	5.03	4.97	7.59	-	5.16	6.12
I	5.56	8.47	7.59	25.00	7.59	9.52	11.36	5.16	-	7.41
J	6.67	5.22	6.20	11.00	5.77	7.19	13.48	6.12	7.41	-

In the Hamming metric, the total length of the vector is taken as the denominator (Equation (14)), so the small value of the number of effective bands between two events E_{11} does not affect the response. Based on the results of Table 4, the maximum mismatch between train and test results is 8.1%, which occurred in the drilling and gunshot events. The average difference in all events is 5.7%, which shows that the change in the selected bands based on train and test data is very slight, and only 10 bands (out of 173 bands) differ.

The average Dice difference obtained in the training and testing samples is 4.07% (Table 5), which reflects the good alignment of the selected bands using the training data and the testing data. The maximum difference in this metric is 14.3%, which occurred between the dog bark and siren events. Similar to the Jaccard metric, the reason for this high difference is the low number of effective bands in this case. According to Equation (15), if there are only a few effective bands available, E_{11} is a small value, and a small mismatch between the two vectors causes a large difference. To solve this problem, it is sufficient to increase the number of effective bands by reducing the specified rejection percentage of the mean equality test.

Table 4. Hamming distance (%) between the results of train and test data (the listed events are: A = air conditioner, B = street music, C = children playing, D = dog bark, E = drilling, F = engine idling, G = gun shot, H = jackhammert, I = siren, J = street music).

Class	A	B	C	D	E	F	G	H	I	J
A	-	4.62	5.20	4.62	4.62	4.62	6.36	7.51	5.20	5.78
B	4.62	-	5.78	4.62	4.62	6.94	6.36	5.20	5.78	4.05
C	5.20	5.78	-	4.62	8.09	6.94	4.62	4.62	6.36	4.62
D	4.62	4.62	4.62	-	6.94	5.20	5.20	5.78	5.78	6.36
E	4.62	4.62	8.09	6.94	-	6.36	8.09	4.62	6.36	5.20
F	4.62	6.94	6.94	5.20	6.36	-	5.78	4.62	8.09	6.36
G	6.36	6.36	4.62	5.20	8.09	5.78	-	6.36	5.78	6.94
H	7.51	5.20	4.62	5.78	4.62	4.62	6.36	-	4.62	5.20
I	5.20	5.78	6.36	5.78	6.36	8.09	5.78	4.62	-	5.78
J	5.78	4.05	4.62	6.36	5.20	6.36	6.94	5.20	5.78	-

Table 5. Dice distance (%) between the results of train and test data (the listed events are: A = air conditioner, B = street music, C = children playing, D = dog bark, E = drilling, F = engine idling, G = gun shot, H = jackhammert, I = siren, J = street music).

Class	A	B	C	D	E	F	G	H	I	J
A	-	2.63	3.32	3.03	2.52	2.74	4.00	4.23	2.86	3.45
B	2.63	-	3.91	3.85	2.78	4.23	6.15	3.11	4.42	2.68
C	3.32	3.91	-	3.39	4.67	4.00	3.20	2.68	3.94	3.20
D	3.03	3.85	3.39	-	4.69	2.99	7.32	3.36	14.29	5.82
E	2.52	2.78	4.67	4.69	-	3.49	5.22	2.58	3.94	2.97
F	2.74	4.23	4.00	2.99	3.49	-	3.55	2.55	5.00	3.73
G	4.00	6.15	3.20	7.32	5.22	3.55	-	3.94	6.02	7.23
H	4.23	3.11	2.68	3.36	2.58	2.55	3.94	-	2.65	3.16
I	2.86	4.42	3.94	14.29	3.94	5.00	6.02	2.65	-	3.85
J	3.45	2.68	3.20	5.82	2.97	3.73	7.23	3.16	3.85	-

It can be concluded that when selecting effective frequency bands using training data (Table 2), good results on test data could be achieved, demonstrating the effectiveness of the proposed effective frequency band selection method. Using these tables makes it possible to select effective spectrogram bands for AED systems. Therefore, the proposed method can be considered a suitable scheme for feature selection in AED or classification systems because the rejection percentage is not affected by the feature type.

5. Conclusions and Future Work

In this article, a statistical method for feature analysis was proposed. The proposed method considers each feature value as a statistical population. The samples of each feature are divided into two populations according to belonging or not belonging to a particular class. The means of these two populations are compared using the two-sample t-test. The feature is useful and otherwise useless if the rejection percentage of the mean equality test for these two populations is sufficiently large. To demonstrate the efficiency of this approach, different frequency bands of the acoustic signal spectrogram were analyzed in an AED system. Since the populations in the two-sample t-tests must be expected, various normality tests were performed, and the normality of the spectrogram features was validated. After the normality test, the two-sample t-test was used to analyze the mean equality between all the frequency bands of the spectrogram for every two acoustic events. According to the results, many spectrogram features (approximately 26.3%) could be omitted during the AED design. In this way, in addition to reducing noise, complexity and training time, the number of samples required to train the system is reduced. Moreover, the training and testing sets were analyzed separately, and the results showed an average

difference of 7.77% in the Jaccard, 4.07% in the Dice, and 5.7% in the Hamming metrics. These small values indicate the validity of the obtained results for the test set.

The assumption of normality in the input data is the only limitation of the proposed method. As to future work, the proposed method can be applied to different AED systems, and its efficiency can be evaluated. Further analysis is needed to show the selected frequency bands are as effective as all frequency bands in machine learning or deep learning models. In this case, the proposed approach should be applied to state-of-the-art AED systems and the accuracy of the system with two inputs, i.e., selected bands and all bands, should be compared.

References

1. Hajihashemi, V.; Alavigharahbagh, A.; Oliveira, H.S.; Cruz, P.M.; Tavares, J.M.R. Novel Time-Frequency Based Scheme for Detecting Sound Events from Sound Background in Audio Segments. In Proceedings of the Iberoamerican Congress on Pattern Recognition, Porto, Portugal, 10–13 May 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 402–416. [CrossRef]
2. Hajihashemi, V.; Gharahbagh, A.A.; Cruz, P.M.; Ferreira, M.C.; Machado, J.J.; Tavares, J.M.R. Binaural Acoustic Scene Classification Using Wavelet Scattering, Parallel Ensemble Classifiers and Nonlinear Fusion. *Sensors* **2022**, *22*, 1535. [CrossRef] [PubMed]
3. Roy, E.; Montrésor, S.; Abraham, P.; Saumet, J.L. Spectrogram analysis of arterial Doppler signals for off-line automated HITS detection. *Ultrasound Med. Biol.* **1999**, *25*, 349–359. [CrossRef] [PubMed]
4. Ibs-von Seht, M. Detection and identification of seismic signals recorded at Krakatau volcano (Indonesia) using artificial neural networks. *J. Volcanol. Geotherm. Res.* **2008**, *176*, 448–456. [CrossRef]
5. Hafez, A.G.; Khan, T.A.; Kohda, T. Earthquake onset detection using spectro-ratio on multi-threshold time–frequency sub-band. *Digit. Signal Process.* **2009**, *19*, 118–126. [CrossRef]
6. Broussard, J.I.; Givens, B. Low frequency oscillations in rat posterior parietal cortex are differentially activated by cues and distractors. *Neurobiol. Learn. Mem.* **2010**, *94*, 191–198. [CrossRef] [PubMed]
7. Liu, D.; Yang, X.; Wang, G.; Ma, J.; Liu, Y.; Peng, C.K.; Zhang, J.; Fang, J. HHT based cardiopulmonary coupling analysis for sleep apnea detection. *Sleep Med.* **2012**, *13*, 503–509. [CrossRef] [PubMed]
8. Dennis, J.; Tran, H.D.; Chng, E.S. Overlapping sound event recognition using local spectrogram features and the generalised hough transform. *Pattern Recognit. Lett.* **2013**, *34*, 1085–1093. [CrossRef]
9. Towsey, M.; Wimmer, J.; Williamson, I.; Roe, P. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecol. Inform.* **2014**, *21*, 110–119. [CrossRef]
10. Vales, D.; Dias, N.A.; Rio, I.; Matias, L.; Silveira, G.; Madeira, J.; Weber, M.; Carrilho, F.; Haberland, C. Intraplate seismicity across the Cape Verde swell: A contribution from a temporary seismic network. *Tectonophysics* **2014**, *636*, 325–337. [CrossRef]
11. de Oliveira, A.G.; Ventura, T.M.; Ganchev, T.D.; de Figueiredo, J.M.; Jahn, O.; Marques, M.I.; Schuchmann, K.L. Bird acoustic activity detection based on morphological filtering of the spectrogram. *Appl. Acoust.* **2015**, *98*, 34–42. [CrossRef]
12. Ghosh, R.; Akula, A.; Kumar, S.; Sardana, H. Time–frequency analysis based robust vehicle detection using seismic sensor. *J. Sound Vib.* **2015**, *346*, 424–434. [CrossRef]
13. Xie, Z.; McLoughlin, I.; Zhang, H.; Song, Y.; Xiao, W. A new variance-based approach for discriminative feature extraction in machine hearing classification using spectrogram features. *Digit. Signal Process.* **2016**, *54*, 119–128. [CrossRef]
14. Xie, J.; Michael, T.; Zhang, J.; Roe, P. Detecting frog calling activity based on acoustic event detection and multi-label learning. *Procedia Comput. Sci.* **2016**, *80*, 627–638. [CrossRef]

15. Sánchez-Gendriz, I.; Padovese, L.R. A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas. *Ecol. Inform.* **2017**, *41*, 1–10. [CrossRef]

16. Zhao, Z.; Zhang, S.H.; Xu, Z.y.; Bellisario, K.; Dai, N.H.; Omrani, H.; Pijanowski, B.C. Automated bird acoustic event detection and robust species classification. *Ecol. Inform.* **2017**, *39*, 99–108. [CrossRef]

17. Shervegar, M.V.; Bhat, G.V. Automatic segmentation of phonocardiogram using the occurrence of the cardiac events. *Inform. Med. Unlocked* **2017**, *9*, 6–10. [CrossRef]

18. Noble, D.J.; MacDowell, C.J.; McKinnon, M.L.; Neblett, T.I.; Goolsby, W.N.; Hochman, S. Use of electric field sensors for recording respiration, heart rate, and stereotyped motor behaviors in the rodent home cage. *J. Neurosci. Methods* **2017**, *277*, 88–100. [CrossRef] [PubMed]

19. Ye, J.; Kobayashi, T.; Murakawa, M. Urban sound event classification based on local and global features aggregation. *Appl. Acoust.* **2017**, *117*, 246–256. [CrossRef]

20. Goenka, A.; Boro, A.; Yozawitz, E. Comparative sensitivity of quantitative EEG (QEEG) spectrograms for detecting seizure subtypes. *Seizure* **2018**, *55*, 70–75. [CrossRef]

21. Hoyos-Barceló, C.; Monge-Álvarez, J.; Pervez, Z.; San-José-Revuelta, L.M.; Casaseca-de-la Higuera, P. Efficient computation of image moments for robust cough detection using smartphones. *Comput. Biol. Med.* **2018**, *100*, 176–185. [CrossRef]

22. Waldman, Z.J.; Shimamoto, S.; Song, I.; Orosz, I.; Bragin, A.; Fried, I.; Engel, J., Jr.; Staba, R.; Sperling, M.R.; Weiss, S.A. A method for the topographical identification and quantification of high frequency oscillations in intracranial electroencephalography recordings. *Clin. Neurophysiol.* **2018**, *129*, 308–318. [CrossRef]

23. Yan, P.Z.; Wang, F.; Kwok, N.; Allen, B.B.; Keros, S.; Grinspan, Z. Automated spectrographic seizure detection using convolutional neural networks. *Seizure* **2019**, *71*, 124–131. [CrossRef]

24. Oliva, J.T.; Rosa, J.L.G. Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection. *Biomed. Signal Process. Control* **2021**, *66*, 102469. [CrossRef]

25. Zhang, M.; Li, Y.; Chen, J.; Song, Y.; Zhang, J.; Wang, M. Event detection method comparison for distributed acoustic sensors using φ-OTDR. *Opt. Fiber Technol.* **2019**, *52*, 101980. [CrossRef]

26. Sahai, A.; Weber, R.; McWilliams, B. Spectrogram feature losses for music source separation. In Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO), A Coruña, Spain, 2–6 September 2019; pp. 1–5. [CrossRef]

27. Lin, L.; Wang, X.; Liu, H.; Qian, Y. Guided learning convolution system for dcase 2019 task 4. *arXiv* **2019**, arXiv:1909.06178. https://doi.org/10.48550/arXiv.1909.06178.

28. Spadini, T.; Silva, D.L.d.O.; Suyama, R. Sound event recognition in a smart city surveillance context. *arXiv* **2019**, arXiv:1910.12369. https://doi.org/10.48550/arXiv.1910.12369.

29. Su, Y.; Zhang, K.; Wang, J.; Madani, K. Environment sound classification using a two-stream CNN based on decision-level fusion. *Sensors* **2019**, *19*, 1733. [CrossRef]

30. Gloaguen, J.R.; Can, A.; Lagrange, M.; Petiot, J.F. Road traffic sound level estimation from realistic urban sound mixtures by Non-negative Matrix Factorization. *Appl. Acoust.* **2019**, *143*, 229–238. [CrossRef]

31. Sattar, F.; Driessen, P.; Tzanetakis, G.; Page, W. A new event detection method for noisy hydrophone data. *Appl. Acoust.* **2020**, *159*, 107056. [CrossRef]

32. Lapins, S.; Roman, D.C.; Rougier, J.; De Angelis, S.; Cashman, K.V.; Kendall, J.M. An examination of the continuous wavelet transform for volcano-seismic spectral analysis. *J. Volcanol. Geotherm. Res.* **2020**, *389*, 106728. [CrossRef]

33. Vafeiadis, A.; Votis, K.; Giakoumis, D.; Tzovaras, D.; Chen, L.; Hamzaoui, R. Audio content analysis for unobtrusive event detection in smart homes. *Eng. Appl. Artif. Intell.* **2020**, *89*, 103226. [CrossRef]

34. Znidersic, E.; Towsey, M.; Roy, W.K.; Darling, S.E.; Truskinger, A.; Roe, P.; Watson, D.M. Using visualization and machine learning methods to monitor low detectability species—The least bittern as a case study. *Ecol. Inform.* **2020**, *55*, 101014. [CrossRef]

35. Robinet, F.; Arnaud, N.; Leroy, N.; Lundgren, A.; Macleod, D.; McIver, J. Omicron: A tool to characterize transient noise in gravitational-wave detectors. *SoftwareX* **2020**, *12*, 100620. [CrossRef]

36. Azab, A.; Khasawneh, M. Msic: Malware spectrogram image classification. *IEEE Access* **2020**, *8*, 102007–102021. [CrossRef]

37. Kacha, A.; Grenez, F.; Orozco-Arroyave, J.R.; Schoentgen, J. Principal component analysis of the spectrogram of the speech signal: interpretation and application to dysarthric speech. *Comput. Speech Lang.* **2020**, *59*, 114–122. [CrossRef]

38. Zeng, Z.; Amin, M.G.; Shan, T. Arm motion classification using time-series analysis of the spectrogram frequency envelopes. *Remote Sens.* **2020**, *12*, 454. [CrossRef]

39. Franzoni, V.; Biondi, G.; Milani, A. Emotional sounds of crowds: Spectrogram-based analysis using deep learning. *Multimed. Tools Appl.* **2020**, *79*, 36063–36075. [CrossRef]

40. Sinha, H.; Awasthi, V.; Ajmera, P.K. Audio classification using braided convolutional neural networks. *IET Signal Process.* **2020**, *14*, 448–454. [CrossRef]

41. Luz, J.S.; Oliveira, M.C.; Araujo, F.H.; Magalhães, D.M. Ensemble of handcrafted and deep features for urban sound classification. *Appl. Acoust.* **2021**, *175*, 107819. [CrossRef]

42. Gupta, V.; Mittal, M.; Mittal, V.; Gupta, A. ECG signal analysis using CWT, spectrogram and autoregressive technique. *Iran J. Comput. Sci.* **2021**, *4*, 265–280. [CrossRef]

43. Manhertz, G.; Bereczky, A. STFT spectrogram based hybrid evaluation method for rotating machine transient vibration analysis. *Mech. Syst. Signal Process.* **2021**, *154*, 107583. [CrossRef]

44. Lara, F.; Lara-Cueva, R.; Larco, J.C.; Carrera, E.V.; León, R. A deep learning approach for automatic recognition of seismo-volcanic events at the Cotopaxi volcano. *J. Volcanol. Geotherm. Res.* **2021**, *409*, 107142. [CrossRef]

45. Pham, L.; Phan, H.; Nguyen, T.; Palaniappan, R.; Mertins, A.; McLoughlin, I. Robust acoustic scene classification using a multi-spectrogram encoder-decoder framework. *Digit. Signal Process.* **2021**, *110*, 102943. [CrossRef]

46. Liu, F.; Shen, T.; Luo, Z.; Zhao, D.; Guo, S. Underwater target recognition using convolutional recurrent neural networks with 3-D Mel-spectrogram and data augmentation. *Appl. Acoust.* **2021**, *178*, 107989. [CrossRef]

47. Kadyan, V.; Bawa, P. Transfer learning through perturbation-based in-domain spectrogram augmentation for adult speech recognition. *Neural Comput. Appl.* **2022**, *34*, 21015–21033. [CrossRef]

48. Pahuja, R.; Kumar, A. Sound-spectrogram based automatic bird species recognition using MLP classifier. *Appl. Acoust.* **2021**, *180*, 108077. [CrossRef]

49. Zhang, T.; Feng, G.; Liang, J.; An, T. Acoustic scene classification based on Mel spectrogram decomposition and model merging. *Appl. Acoust.* **2021**, *182*, 108258. [CrossRef]

50. Cheng, K.W.; Chow, H.M.; Li, S.Y.; Tsang, T.W.; Ng, H.L.B.; Hui, C.H.; Lee, Y.H.; Cheng, K.W.; Cheung, S.C.; Lee, C.K.; et al. Spectrogram-based classification on vehicles with modified loud exhausts via convolutional neural networks. *Appl. Acoust.* **2023**, *205*, 109254. [CrossRef]

51. Wang, X.; Jiang, J.; Duan, F.; Liang, C.; Li, C.; Sun, Z.; Lu, R.; Li, F.; Xu, J.; Fu, X. A method for enhancement and automated extraction and tracing of Odontoceti whistle signals base on time-frequency spectrogram. *Appl. Acoust.* **2021**, *176*, 107698. [CrossRef]

52. You, L.; Coyotl, E.P.; Gunturu, S.; Van Segbroeck, M. Transformer-Based Bioacoustic Sound Event Detection on Few-Shot Learning Tasks. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.

53. Bhangale, K.B.; Kothandaraman, M. Speech emotion recognition using the novel PEmoNet (Parallel Emotion Network). *Appl. Acoust.* **2023**, *212*, 109613. [CrossRef]

54. Özseven, T. Investigation of the effectiveness of time-frequency domain images and acoustic features in urban sound classification. *Appl. Acoust.* **2023**, *211*, 109564. [CrossRef]

55. Latif, S.; Shahid, A.; Qadir, J. Generative emotional AI for speech emotion recognition: The case for synthetic emotional speech augmentation. *Appl. Acoust.* **2023**, *210*, 109425. [CrossRef]

56. Shafik, A.; Sedik, A.; Abd El-Rahiem, B.; El-Rabaie, E.S.M.; El Banby, G.M.; Abd El-Samie, F.E.; Khalaf, A.A.; Song, O.Y.; Iliyasu, A.M. Speaker identification based on Radon transform and CNNs in the presence of different types of interference for Robotic Applications. *Appl. Acoust.* **2021**, *177*, 107665. [CrossRef]

57. Mushtaq, Z.; Su, S.F.; Tran, Q.V. Spectral images based environmental sound classification using CNN with meaningful data augmentation. *Appl. Acoust.* **2021**, *172*, 107581. [CrossRef]

58. Sharan, R.V.; Berkovsky, S.; Navarro, D.F.; Xiong, H.; Jaffe, A. Detecting pertussis in the pediatric population using respiratory sound events and CNN. *Biomed. Signal Process. Control* **2021**, *68*, 102722. [CrossRef]

59. Haleem, M.S.; Castaldo, R.; Pagliara, S.M.; Petretta, M.; Salvatore, M.; Franzese, M.; Pecchia, L. Time adaptive ECG driven cardiovascular disease detector. *Biomed. Signal Process. Control* **2021**, *70*, 102968. [CrossRef]

60. Wei, L.; Ventura, S.; Ryan, M.A.; Mathieson, S.; Boylan, G.B.; Lowery, M.; Mooney, C. Deep-spindle: An automated sleep spindle detection system for analysis of infant sleep spindles. *Comput. Biol. Med.* **2022**, *150*, 106096. [CrossRef] [PubMed]

61. Nadalin, J.K.; Eden, U.T.; Han, X.; Richardson, R.M.; Chu, C.J.; Kramer, M.A. Application of a convolutional neural network for fully-automated detection of spike ripples in the scalp electroencephalogram. *J. Neurosci. Methods* **2021**, *360*, 109239. [CrossRef]

62. Wu, Y.; Pang, X.; Zhao, G.; Yue, H.; Lei, W.; Wang, Y. A novel approach to diagnose sleep apnea using enhanced frequency extraction network. *Comput. Methods Programs Biomed.* **2021**, *206*, 106119. [CrossRef] [PubMed]

63. Xie, J.; Aubert, X.; Long, X.; van Dijk, J.; Arsenali, B.; Fonseca, P.; Overeem, S. Audio-based snore detection using deep neural networks. *Comput. Methods Programs Biomed.* **2021**, *200*, 105917. [CrossRef]

64. Wodecki, J.; Michalak, A.; Zimroz, R. Local damage detection based on vibration data analysis in the presence of Gaussian and heavy-tailed impulsive noise. *Measurement* **2021**, *169*, 108400. [CrossRef]

65. Patil, S.; Wani, K. Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network. *Mater. Today Proc.* **2023**, *72*, 1322–1327. [CrossRef]

66. Wu, H.; Huang, A.; Sutherland, J.W. Condition-Based Monitoring and Novel Fault Detection Based on Incremental Learning Applied to Rotary Systems. *Procedia CIRP* **2022**, *105*, 788–793. [CrossRef]

67. Kahl, S.; Wood, C.M.; Eibl, M.; Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.* **2021**, *61*, 101236. [CrossRef]

68. Katsis, L.K.; Hill, A.P.; Piña-Covarrubias, E.; Prince, P.; Rogers, A.; Doncaster, C.P.; Snaddon, J.L. Automated detection of gunshots in tropical forests using convolutional neural networks. *Ecol. Indic.* **2022**, *141*, 109128. [CrossRef]

69. Trani, L.; Pagani, G.A.; Zanetti, J.P.P.; Chapeland, C.; Evers, L. DeepQuake—An application of CNN for seismo-acoustic event classification in The Netherlands. *Comput. Geosci.* **2022**, *159*, 104980. [CrossRef]

70. Meng, J.; Wang, X.; Wang, J.; Teng, X.; Xu, Y. A capsule network with pixel-based attention and BGRU for sound event detection. *Digit. Signal Process.* **2022**, *123*, 103434. [CrossRef]

71. Harris, F.J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **1978**, *66*, 51–83. [CrossRef]

72. Miller, S.; Childers, D. *Probability and Random Processes: With Applications to Signal Processing and Communications*; Academic Press: Cambridge, MA, USA, 2012. [CrossRef]

73. Rozanov, Y. *Probability Theory, Random Processes and Mathematical Statistics*; Springer Science & Business Media: Dordrecht, The Netherlands, 2012; Volume 344. [CrossRef]

74. Hatem, G.; Zeidan, J.; Goossens, M.; Moreira, C. Normality testing methods and the importance of skewness and kurtosis in statistical analysis. *BAU J.-Sci. Technol.* **2022**, *3*, 7. [CrossRef]

75. Radhi, A.A.; Abdullah, H.N.; Akkar, H.A. Denoised Jarque-Bera features-based K-Means algorithm for intelligent cooperative spectrum sensing. *Digit. Signal Process.* **2022**, *129*, 103659. [CrossRef]

76. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]

77. Royston, J. Some techniques for assessing multivarate normality based on the Shapiro-Wilk W. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1983**, *32*, 121–133. [CrossRef]

78. Sarhan, A.E.; Greenberg, B.G. Estimation of location and scale parameters by order statistics from singly and doubly censored samples. *Ann. Math. Stat.* **1956**, *27*, 427–451. [CrossRef]

79. Villasenor Alva, J.A.; Estrada, E.G. A generalization of Shapiro–Wilk's test for multivariate normality. *Commun. Stat.-Theory Methods* **2009**, *38*, 1870–1883. [CrossRef]

80. Yazici, B.; Yolacan, S. A comparison of various tests of normality. *J. Stat. Comput. Simul.* **2007**, *77*, 175–183. [CrossRef]

81. Uhm, T.; Yi, S. A comparison of normality testing methods by empirical power and distribution of P-values. *Commun. Stat.-Simul. Comput.* **2023**, *52*, 4445–4458. [CrossRef]

82. Seier, E. Comparison of tests for univariate normality. *InterStat Stat. J.* **2002**, *1*, 1–17.

83. Eelbode, T.; Bertels, J.; Berman, M.; Vandermeulen, D.; Maes, F.; Bisschops, R.; Blaschko, M.B. Optimization for medical image segmentation: Theory and practice when evaluating with dice score or jaccard index. *IEEE Trans. Med. Imaging* **2020**, *39*, 3679–3690. [CrossRef]

84. Costa, L.D.F. On similarity. *Phys. A Stat. Mech. Its Appl.* **2022**, *599*, 127456. [CrossRef]

85. Adzhemov, A.; Kudryashova, A. Features of Converting Signals to Binary and Minimizing Distortion. In Proceedings of the 2021 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 16–18 March 2021; pp. 1–5.

86. Yadav, E.; Chawla, V. Fault detection in rotating elements by using fuzzy integrated improved local binary pattern method. *J. Braz. Soc. Mech. Sci. Eng.* **2022**, *44*, 596. [CrossRef]

87. Razali, N.M.; Wah, Y.B. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J. Stat. Model. Anal.* **2011**, *2*, 21–33.