applied sciences

MDPI

Article

An Audio-Based Method for Assessing Proper Usage of Dry Powder Inhalers

Athina-Chara Eleftheriadou, Anastasios Vafeiadis, Antonios Lalas *, Konstantinos Votis and Dimitrios Tzovaras

Centre for Research and Technology Hellas, Information Technologies Institute, 6th km Charilaou-Thermi, 57001 Thermi, Greece; athielef@iti.gr (A.-C.E.); anasvaf@iti.gr (A.V.); kvotis@iti.gr (K.V.); dimitrios.tzovaras@iti.gr (D.T.)

* Correspondence: lalas@iti.gr

Received: 31 August 2020; Accepted: 21 September 2020; Published: 24 September 2020

Abstract: Critical technique errors are very often performed by patients in the use of Dry Powder Inhalers (DPIs) resulting in a reduction of the clinical efficiency of such medication. Those critical errors include: pure inhalation, non-arming of the device, no exhalation before or after inhalation, and non-holding of breath for 5–10 s between inhalation and exhalation. In this work, an audio-based classification method that assesses patient DPI user technique is presented by extracting the the non-silent audio segments and categorizing them into respiratory sounds. Twenty healthy and non-healthy volunteers used the same placebo inhaler (Bretaris Genuair Inhaler) in order to evaluate the performance of the algorithm. The audio-based method achieved an F1-score of 89.87% in classifying sound events (*Actuation*, *Inhale*, *Button Press*, and *Exhale*). The significance of the algorithm lies not just on automatic classification but on a post-processing step of peak detection that resulted in an improvement of 5.58% on the F1-score, reaching 94.85%. This method can provide a clinically accurate assessment of the patient's inhaler use without the supervision of a doctor.

Keywords: audio classification; machine learning; feature extraction; MFCCs; asthma; COPD; DPIs

1. Introduction

Asthma and Chronic Obstructive Pulmonary Disease (COPD) are pulmonary diseases that have affected millions of people worldwide. Nowadays, it is estimated that 300 million people are suffering from asthma worldwide [1]. Asthma is considered to be the most common non-contagious disease among children, but is still a disease affecting adults during their lifetime and most of the deaths occur in older adults [2]. At the same time, COPD is considered one of the major causes of chronic morbidity while being the fourth leading cause of death in the world [3] and is expected to be the third by the end of 2020 [4]. Such inflammatory lung diseases not only significantly downgrade the quality of patients' and their families' lives but are also deteriorate the efficiency of the healthcare system [5].

Asthma and COPD are treated using handheld inhalers as prior medication. Inhalers deliver medication directly to the airways. Critical errors, which are common to occur, can reduce the medication amount delivered to the patient's lungs, or even annihilate it. Pure inhalation, non-arming of the device, no exhalation before or after inhalation, and non-holding of breath for 5–10 s between inhalation and exhalation are the main errors listed [6–8]. Such errors result in poor disease control, increased medical care, and high risks of mortality [9]. The costs because of wasted medication due to poor inhalation technique are between $5 and $7 billion [10], while poor inhalation technique was estimated to be associated with costs of €782 million across Spain, Sweden, and the UK in 2015 [11]. The main method to assess a patient's inhaler user technique is by using checklists based

on visual/aural assessment, where a healthcare professional supervises the whole process [11,12]. However, this method of assessment is based on the healthcare professional's perception, it gives an equal rating to all errors, it could overestimate patient performance and it cannot be used to monitor how patients use their inhaler outside of clinical visits, without the supervision of a doctor [13,14]. It is reported that the inhaler user technique from patients can be improved by inhaler training devices [15]. Guiding patients to improve their inhaler usage technique while providing proper feedback to medical personnel could avoid dangerous exacerbation events while facilitating effective self-management of obstructive respiratory diseases [16,17]. This is the first step in guiding the patients by evaluating the followed procedure from the audio information given through either a smartphone's microphone or through an attachable microphone that is connected to the inhaler.

A DPI is a device that disperses a dry powder medication to the lungs. The medication is released only when the patient takes a deep, strong breath in through the inhaler; therefore, it is breath-activated. The difference between Dry Powder Inhalers (DPIs) and Metered Dose Inhalers (MDIs) is that MDIs push medication into the lungs. Dry powder inhalers are the third type of inhaler device, after MDIs and nebulizers, and among these seems to be the most promising devices for future use [18]. Such devices carry the dose into the lungs (actuation) when breathing is performed by the patient and this way there is no need for coordination of patient's breathing and the activation of the device. Furthermore, DPIs have other important attributes that make them preferable among other devices [19]. Promising physical stability associated with DPIs is due to the solid form of the formulation.

Proper inhaler use includes the following steps (details can be found in [20]) (shown, also, in Figure 1): remove the cap; hold inhaler horizontally with the colored button facing up; press and release the button; breathe out fully (away from inhaler); breathe in strongly and deeply with mouthpiece in mouth and the lips sealed; hold breath for 5 to 10 s; breathe out gently (away from inhaler); and replace cap. Despite these attributes, as it has been aforementioned, patients often make critical errors, such as pure inhalation, non-arming of the device, no exhalation before or after inhalation and non-holding of the breath for more than 5 s. Those errors prohibit them from receiving full therapeutic effect from their medication.

For the scope of this study, the Bretaris Genuair placebo Inhaler [20] was used for the data capture procedure. Genuair is a multi-dose inhaler, giving the advantage to patients of carrying many doses wherever they go. Genuair is, also, reusable (in contrast with Diskus and Turbuhaler [21]), being cost effective over time. Moreover, Negro et al. [22] tested the usability of the seven most used DPIs in COPD and Genuair ranked sixth in a scale of usability, indicating a high factor of errors to be performed during its usage. Mainly, Genuair was used for the purpose of this study because it was stated that it ranks in a high place in the Greek market of inhaler devices; thus, it is very common among respiratory patients. All the above attributes of Genuair triggered our research towards the device utilization and assessment.

Detection and classification of sound events produced by the use of an inhaler is a field with emerging research as is shown by recent work [13,23]. Holmes et al. [24] developed audio-signal processing methods to automatically detect inhaler events. An INhaler Compliance Assessment (INCA) device, which uses a microphone with sampling rate of 8 kHz and depth of 8 bits/sample, was used to obtain audio recordings of Diskus DPI use from 12 asthma patients. The study obtained 609 audio recordings. A set of features were employed to characterize the different inhaler events, such as the power of specific frequency bands, Mel-Frequency Cepstral Coefficients (MFCCs) and three types of sounds were classified using feature value decision thresholds: preparation (blister), inhalation, and exhalation. The algorithm achieved an accuracy of 92.1% for preparation detection, 91.7% for inhalation detection, and 93.7% for exhalation detection. Taylor et al. [25] developed an accurate method of detecting Metered Dose Inhaler (MDI) actuations using inhaler audio recordings (sampled at 44.1 kHz, with depth of 16 bits/sample) obtained from five healthy participants and 15 hospitalized asthma and COPD patients. There were 179 audio signals obtained using an attachable

microphone fixed to the side of a pMDI. This method accurately distinguished actuations from other inhaler events using the continuous wavelet transform (by analyzing the high frequency power content of the signal). The method was trained on 20 and tested on 159 audio signals with an accuracy of 99.7%, sensitivity of 100%, and specificity of 99.4% at detecting pMDI actuations.

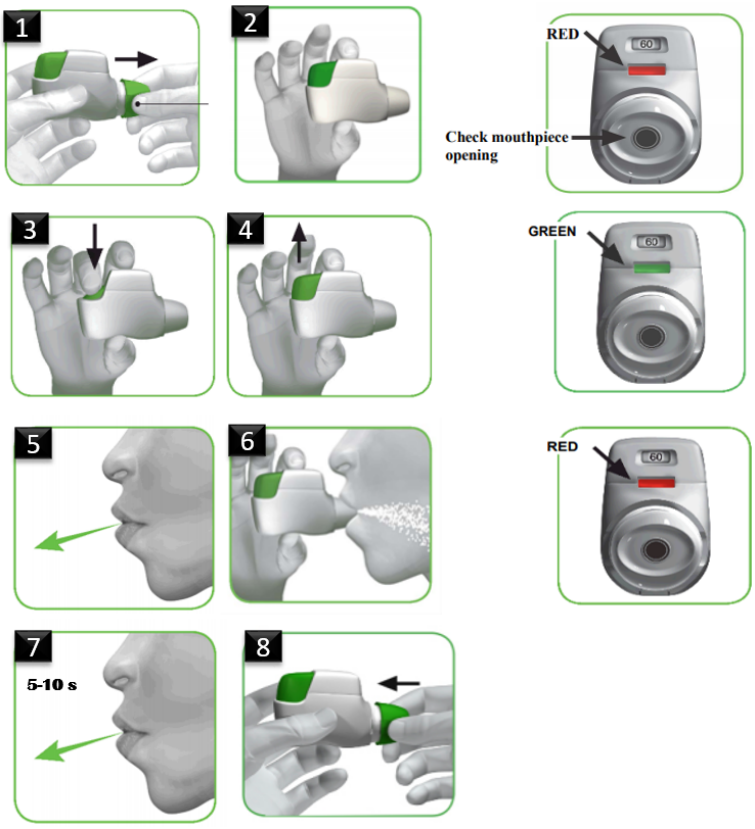

Figure 1. Dry Powder Inhaler (DPI) usage description as in [20].

Kikidis et al. [26] investigated the use of convolution neural networks (CNNs) to detect pMDI actuation sounds. The authors used a microphone (connected to a smartphone) that was fixed onto the front side of a pMDI. The pMDI audio signals were acquired from five healthy participants who were asked to trigger the pMDI in open air and away from their mouth in different real-life noisy environments. This way they produced 400 recordings (200 of actuation and 200 of non-actuation) and an accuracy of 98% was reported at detecting pMDI actuation sounds. Nousias et al. [27] obtained pMDI audio data from five healthy participants used for the classification of four sound types or classes (inhaler actuation, inhalation, exhalation, background noise). The dataset was comprised of 280 samples per sound class giving a total of 1120 sound samples in the dataset (sampled at 4 kHz with depth of 4 bits/sample). In this study, it was reported that AdaBoost outperformed the alternative approaches leading to accuracies above 96% by using the Short Time Fourier Transform (STFT) as the feature extraction method. Lalos et al. [28] developed an energy efficient wireless audio-based method of remotely detecting pMDI events. In this study, a dataset from two healthy participants consisting of 500 actuations and 500 noise segments (dataset 1) along with 200 actuations, 200 inhalations, 200 exhalations, and 200 noise segments (dataset 2) were obtained. An accuracy of over 96% was reported for the two datasets.

In this paper, motivated by the pure research content in terms of identifying and evaluating the proper use of DPIs, a content-based audio classification approach for identifying four different respiratory sounds, i.e., *Actuation*, *Inhale*, *Exhale*, and *Button Press* is proposed. The method is focused on the usage of traditional classifiers, and particularly on the feature extraction. Additionally, segmentation is not performed using a sliding window approach, however, silence removal is

performed to identify the segments containing sound. It is a multiclass classification problem with the two classes (*Actuation* and *Inhale*) having minor differences making it more challenging to distinguish them. This separation between them is the success of the algorithm which is not based just to the automatic results from the prediction model but also to a post-processing step that is being included. A new dataset was collected within this research's goals which consists of 400 audio recordings per class (1600 in total), generated using 20 subjects. The algorithm achieved near real-time performance, as it is denoted by the recorded inference times.

The paper is organized as follows: Section 2 describes the dataset and how the audio data files where collected for the algorithms training, the data pre-processing step, the procedure of segmenting the acoustic signal into non-silent intervals, as well as the feature extraction procedure. The classification results, the post-processing step that has been applied, and the inference times are presented in Section 3. Finally, the conclusion points are analyzed in Section 4.

2. Methodology

2.1. Data Collection

Various audio signals were collected in an ambient environment but also in a noisy environment—indoors and outdoors. Twenty healthy and asthmatic volunteers, aged 18–80 years, used the same placebo inhaler, Genuair Inhaler, Menarini Hellas [20]. The sounds were recorded using different smartphones since each smartphone has a different type of microphone, with microphones having different frequency responses, and by default different suppression algorithms. This approach was applied in order to make the algorithm easy to generalize and as robust as possible. In this context, it covers a wide range of different microphones to provide, also, extensive usability for different patients' mobile devices. A total of 1600 audio files were collected for all classes. Each file has a duration of 2 s, with a sampling rate of 8 kHz and a 16-bit depth. The signals were categorized into four classes (Figure 2): *Actuation*, *Inhale*, *Button Press*, and *Exhale*.

- *Actuation*: inhalation using the device when hearing the "click" and indicating the correct intake of the drug.
- *Inhale*: inhalation using the device without releasing the drug (absence of the "click" sound) either due to pure inhalation or non-arming of the device and indicating the wrong intake of the drug.
- *Button Press*: sound produced when button is pressed.
- *Exhale*: exhaling of the patient away from the inhaler.

Data Pre-Processing

It was observed that each sound event can be completed within an average time frame of 2 s. *Button Press* is about 500 ms long, *Actuation* is between 1 to 3 s, *Inhale* is about 2 s, while *Exhale* varies from 1.5 to 4 s. To train the classifier with the audio-extracted features, the audio samples must be of the same length. For this purpose audio files that were less than 2 s were zero padded at the end of the signal, while the larger ones were trimmed to 2 s without losing important signal information.

In addition, during the signal pre-processing stage, the files to be evaluated are passed through a high-pass filter. The high-pass filter is used to clear low-frequency noise and eliminate low-frequency voltages (DC offset removal). The filter selected is a high-pass Butterworth filter [29], which is designed to have a frequency response as flat as possible in the transit zone (i.e., no ripples) and slides to zero in the cut-off zone. Therefore, the filter with a cut-off frequency of 20 Hz and order = 3 was applied.

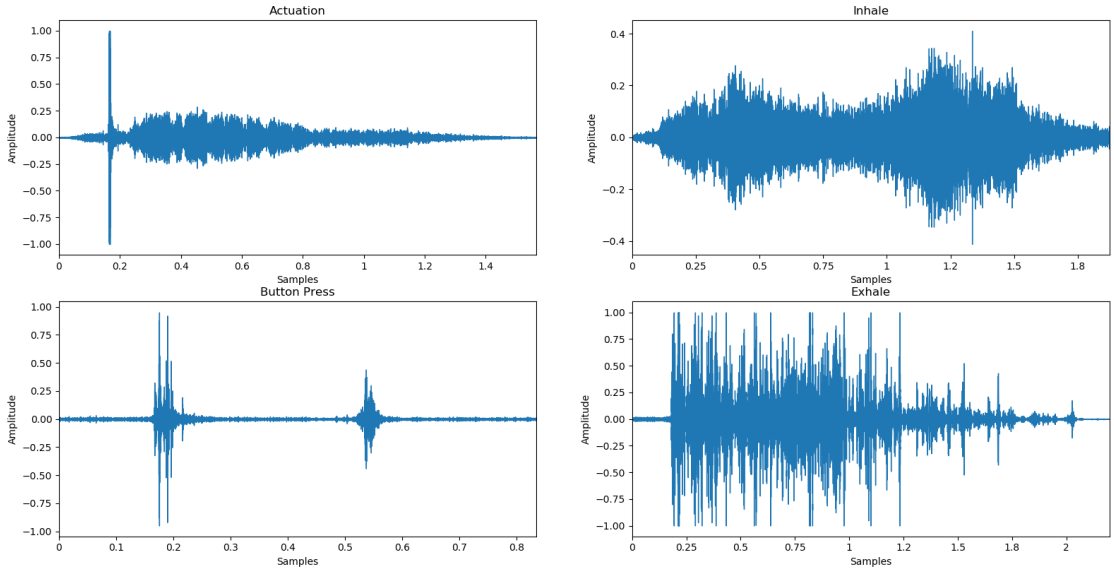

Figure 2. Waveforms of audio signals corresponding to the four classes, i.e., *Actuation*, *Inhale*, *Exhale*, and *Button Press*.

2.2. Silence Removal

In the process of extracting segments that contain respiratory sounds from the acoustic signal, the algorithm that is presented takes the input signal and returns the endpoints of the segments corresponding to the individual sound events. This process belongs to the broader field of voice activity detection, a technique used in speech processing, in which the presence or absence of human speech is detected. The algorithm removes all "silent" areas of the signal and returns those containing sound to classify each of them into the final classes of *Actuation*, *Inhale*, *Exhale*, *Button Press*. For this purpose pyAudioAnalysis [30] library was used.

The first step in the activity detection process is the extraction of 34 short-term features (Table 1). Then, a binary Support Vector Machine (SVM) model is trained to distinguish high-energy from low-energy short-term frames using 50% of the components with the highest energy concentration along with the 50% of the lower ones. The short-term features use, for this purpose, 50 ms window size and 40 ms frame step, resulting in a sequence of feature vectors with 34 elements each. In the next step, the SVM model is trained to distinguish the high-energy short-term windows from the low-energy ones and applies them throughout the recording, using a median threshold value to detect the segments that contain sound. A dynamic threshold is used to detect active segments. Finally, the "non-silent" sections are grouped and the "silent" areas are removed.

Table 1. Audio features extracted in silence removal [30].

No	Feature	Description
1	Zero Crossing Rate	The rate of sign-changes of the signal during the duration of a particular frame.
2	Energy	The sum of the squares of the signal, normalized by the length of the frame.
3	Entropy of Energy	The entropy of the normalized sub-frames.
4	Spectral Centroid	The center of gravity of the spectrum.
5	Spectral Spread	The second central moment of the spectrum.
6	Spectral Entropy	The entropy of normalized spectral energies for a set of sub-frames.
7	Spectral Flux	The square difference between the normalized magnitudes of the spectrum of two consecutive frames.
8	Spectral Roll-off	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.
9–21	MFCCs	MFCCs form a cepstral representation where the frequency bands are not linear but distributed on a mel scale.
22–33	Chroma Vector	A 12 element representation of spectral energy.
34	Chroma Deviation	The standard deviation of the 12 chroma coefficients.

2.3. Feature Extraction

Feature extraction is a very important part of analyzing and finding relationships between audio data. Classification and prediction algorithms require feature extraction. Through a continuous testing process it was deduced that the following six, in number, features are used for the signal analyzing, as discussed below.

2.3.1. Mel Frequency Cepstral Coefficients

The sounds produced by a human are filtered by the shape of the voice path including the tongue, the teeth, and other components which determine which sound comes out. This shape of the voice path is manifested in the "area" of the low-power spectrum and the MFCCs contribute to the representation of the low power spectrum of a sound. The followed procedure is shown in Figure 3 [31].

For the FFT, a window size of 512 frames with number of frames between FFT columns being 128 is used, and 16 MFCCs including the zero coefficient are returned, resulting in an array of size (16×126).

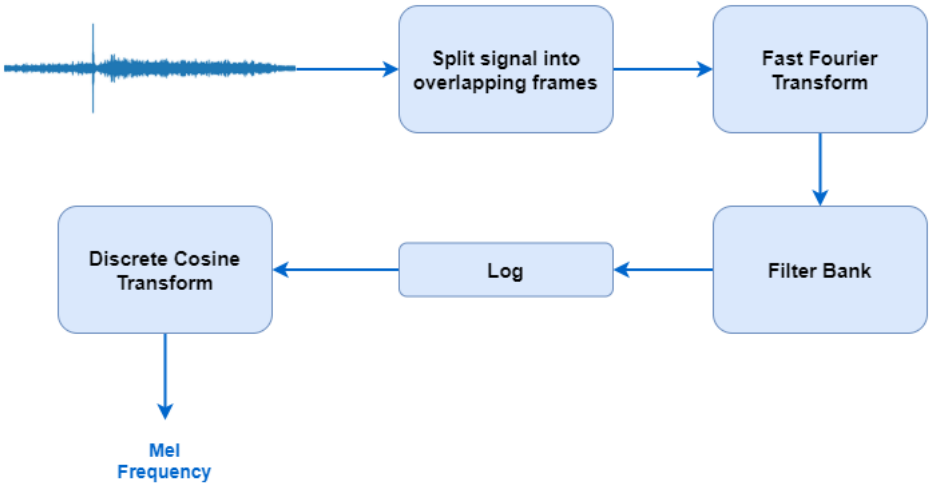

Figure 3. The followed procedure for Mel-Frequency Cepstral Coefficient (MFCC) extraction [31].

2.3.2. Zero Crossing Rate

Zero Crossing Rate (ZCR) is the rate at which the signal changes from positive to zero to negative or from negative to zero to positive [32]. That is, the signal's rate of passage from zero.

ZCR is defined formally as:

$$ZCR = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{R<0}(s_t s_{t-1}) \quad (1)$$

where s is a signal of length T and $1_{R<0}$ is an indicator function.

To compute the ZCR of an audio time series, we set the length of the frame over which to compute the rates to 512 with 128 samples to advance for each frame (hop length), which finally produces an array of size (1×126).

2.3.3. Root Mean Square Energy

The root mean square energy (RMSE) is defined as a measure of the signal intensity.

The followed procedure is:

- The signal is divided into windows.
- For each window, each sample value is squared (multiplied by itself).
- The average is obtained.
- Calculate the square root of the mean.

Given by the formula [33]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n}|x(n)|^2} \tag{2}$$

where $x(n)$ represents the size (amplitude) of the bin number n (bin means dividing the entire range of values into a series of intervals) and N is the number of bins.

To compute the RMSE for each frame, the energy from the audio sample is calculated without the need of a Short Term Fourier Transform, by using a frame length of 512 samples with 128 hop length. An array of size (1×126) is extracted.

2.3.4. Spectral Flatness

Spectral Flatness is defined as the ratio of the geometric mean of the spectrum to the arithmetic mean of the power spectrum and it is a way of measuring how close the audio is to white noise, which has a flat power spectrum (spectral flatness closer to 1.0). The arithmetic mean of a sequence of n elements is what is considered to be an average, add all the elements and divide by n, while the geometric mean is the nth root of the elements' product. It is calculated by taking the arithmetic mean of the logarithms of the objects, and then taking the exponential result [34].

It is given by the formula:

$$Flatness = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} = \frac{exp(\frac{1}{N}\sum_{n=0}^{N-1} ln(x(n)))}{\frac{1}{N}\sum_{n=0}^{N-1} x(n)} \tag{3}$$

where $x(n)$ represents the size of the bin n and N is the number of bins.

To compute spectral flatness, an FFT window size of 512 samples along with hop length of 128 samples were used, resulting in an array of size (1×126).

2.3.5. Spectral Centroid

Spectral Centroid (SC) measures the shape of the signal spectrum, as it is the center of gravity of the spectrum. Centroid is computed by considering the spectrum as a distribution of frequencies (values) and probabilities to observe these frequencies (normalized amplitude). A higher SC value corresponds to more signal energy concentrated at higher frequencies. Therefore, it measures the spectral shape and position of the spectrum [35].

It is given by the formula [36]:

$$SC_t = \frac{\sum_{n=1}^{N} M_t[n]n}{\sum_{n=1}^{N} M_t[n]} \tag{4}$$

where $M_t[n]$ is the magnitude of the Fourier transform at frame t and frequency bin n.

To compute the SC, each frame of a magnitude spectrogram is normalized and the mean (centroid) is extracted per frame using an FFT window size of 512 samples and hop length of 128 samples. An array of size (1×126) is produced.

2.3.6. Spectral Roll-Off

The spectral roll-off is defined for each window as the center frequency for a spectrogram bin, so that this bin and the lower contain at least the roll percentage value (0.85 by default [35]), i.e., the ratio of total energy to attain before yielding the roll-off frequency, of the spectral energy in this frame.

A roll-off frequency R_t such that [37]:

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^{N} M_t[n] \tag{5}$$

where $M_t[n]$ is the magnitude of the Fourier transform at frame t and frequency bin n.

To compute the roll-off frequency, an FFT window size of 512 samples along with hop length of 128 samples were used, resulting in an array of size (1×126).

Finally, all features are concatenated creating an array of size (21×126) for each file of the dataset.

3. Results

3.1. Classification

The last step of this algorithm (as shown in Figure 4), after extracting features, is the process of classifying the audio signals into one of the final classes. The database of 1600 audio files is used to train the classifiers and to find the one that achieves the highest classification performance among the data. These data are split into training and test data, 75% for algorithms' training and 25% for the testing.

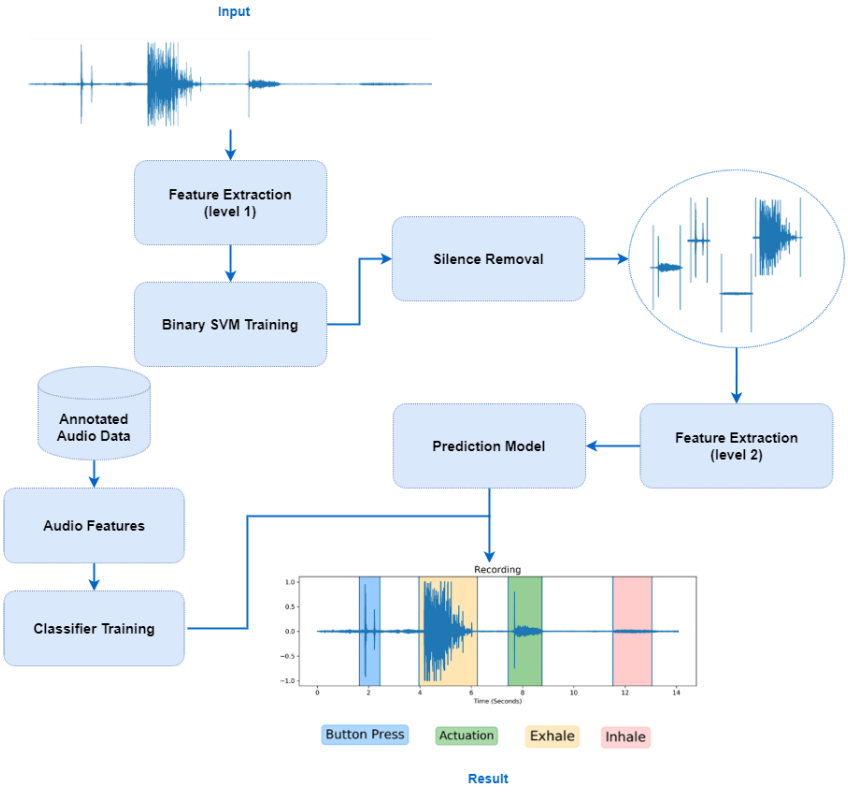

Figure 4. Flow diagram of the proposed algorithm, where first silence removal is performed in the audio signal and then the classification of each extracted segment into four targeted classes, i.e., *Actuation*, *Inhale*, *Button press*, *Exhale*.

Various classifiers were used for this purpose, such as k-Nearest Neighbors (k-NN) [38], Extra Trees [39], SVM [40], Gradient Boosting (GB) [41], and Random Forest [36]. The following metrics were selected to evaluate the results of each classifier: Accuracy, F1-score, precision, and recall. Gradient Boosting achieved the highest accuracy of 90.92% and an F1-score of 89.87% (Table 2). For the performance evaluation of the classifiers, 10-fold cross validation was used and the results were calculated on the testing set. The silence removal procedure is not used in this step since the database consisted of fixed-sized audio files of 2 s.

Gradient Boosting is a technique for classification tasks in the field of machine learning. It produces a prediction model in the form of an ensemble of weak prediction models (decision trees). The model is built in a stage-wise fashion, and the generalization is being performed by an arbitrary differentiable loss function.

Gradient Boosting algorithm can easily overfit [41] a training data set, and that is why different regularization methods could be applied to improve the algorithm's performance and encounter the problem of possible overfitting. A large number of boosting stages to perform usually results in better performance, and for that reason 500 estimators are used. Furthermore, the maximum number of features are taken into account when looking for the best split and the minimum number of samples required to split an internal node are 10. In addition, "learning rate" is set to 0.1 since the parameter's tuning resulted in a better classification score after performing a grid-search between 0.001 to 0.1 with steps of 0.05.

Table 2. Classification results among tested classifiers after 10-fold cross validation.

Classifier	Accuracy	F1-Score	Precision	Recall
Gradient Boosting	90.92%	89.87%	90.44%	*89.91%*
Extra Trees	89.42%	88.40%	88.86%	88.29%
Random Forest	87.22%	86.08%	86.20%	86.20%
SVM	78.96%	77.87%	78.29%	77.87%
k-NN	76.08%	75.63%	77.07%	75.99%

3.2. Post-processing

A post-processing step is used to further improve the algorithm. The performance of the algorithm between the *Actuation* and *Inhale* classes is not as high as the other classes (*Button press* and *Exhale*) (Table 3). This is due to the fact that distinguishing the classes is particularly difficult, as the signals are "similar" to each other. Towards this, observation peak detection is used along with comparison of the maximum peak value with the average signal width to improve the categorization results.

Table 3. Class-wise F1-score using Gradient Boosting (GB) before post processing.

Class	F1-Score
Actuation	87.12%
Inhale	79.35%
Button press	96.78%
Exhale	96.23%

Peak detection is calculated as follows:

- Take as input the segments classified as *Actuation* or *Inhale*.
- Calculate the average signal width.
- Detect peaks in the first 0–8000 samples of the signal.
- Find the maximum peak.

- Perform a comparison between the maximum peak value detected in the range of the samples and the average width of the signal. If (maximum peak—average width) is greater than 90% then the event is classified as *Actuation*. Otherwise as *Inhale*.

This procedure is selected based on the 400 audio files belonging to the *Actuation* class. The analysis of the signal and the peak detection is performed in the first half of the signal (i.e., 0–8000 samples). That is because it was observed that the drug is released at the beginning of the inhalation, while the "click" sound appears in the signal representation as a peak (Figure 5).

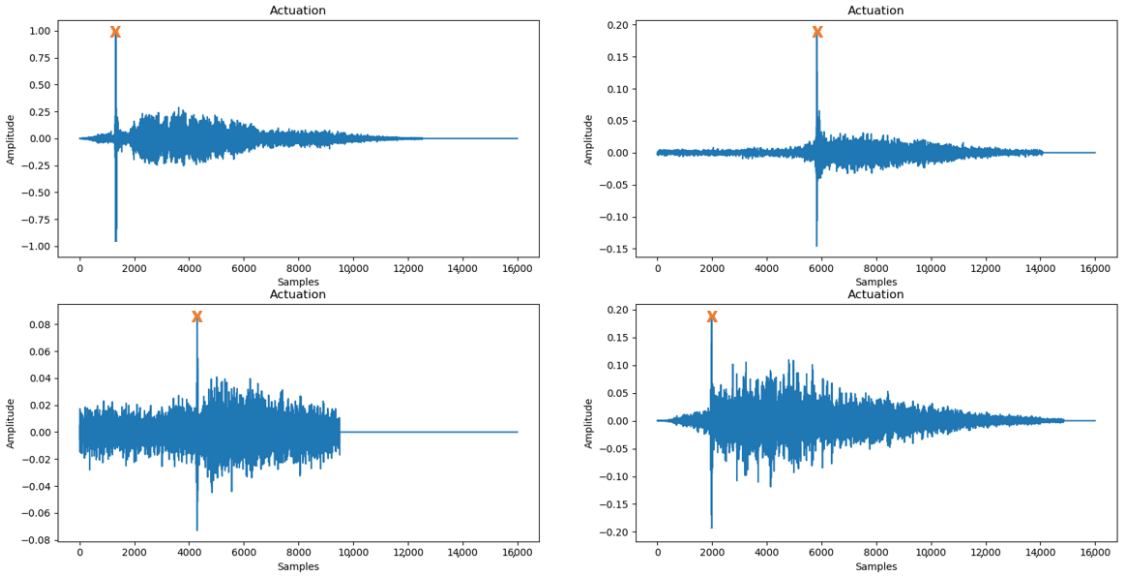

Figure 5. Peak detection (orange *X*) performed successfully in audio files belonging to *Actuation* with different background noise.

Finally, the after-post-processing results are shown in Table 4, where an important improvement of the *Actuation* and *Inhale* classes' F1-score performance is noticed. Table 5 summarizes the metrics extracted from Gradient Boosting's algorithm performance after the post-processing step is performed.

Table 4. F1-score of Gradient Boosting between targeted classes after post-processing.

Class	F1-Score
Actuation	93.12%
Inhale	94.08%
Button press	96.67%
Exhale	95.56%

Table 5. Summarized metrics from the Gradient Boosting Classifier.

Classifier	Accuracy	F1-Score	Precision	Recall
Gradient Boosting	95.21%	94.85%	95.04%	94.67%

3.3. Inference Times

In this section, the inference times of the algorithm are presented for extracting the "non-silent" segments and classifying these segments into one of the target classes. Firstly, the time taken by the Gradient Boosting to classify a single audio file, with a duration of approximately 2 s and containing one respiratory sound, is 25 ms on average. Furthermore, 8 to 14 s audio files are tested, which contain multiple events and require the process of segmentation apart from classification. The algorithm completes the process in approximately 230 ms.

Finally, a larger sound file is introduced, which is approximately 30 s long. Through our interaction with patients, it was denoted that this is the necessary time needed to complete the whole process of the inhaler's use (press button to trigger the device, exhale fully and away from the inhaler, take a deep breath with the inhaler in mouth, and then, when the inhalation is over, hold breath for 5 to 10 s until breathing out again). The entire procedure for classifying this file takes approximately 420 ms.

Despite the complexity and computational cost required by the Gradient Boosting classifier, the process is quickly completed in just a few milliseconds (Figure 6), which, combined with the accuracy of the prediction, makes the algorithm ideal for real-time event detection.

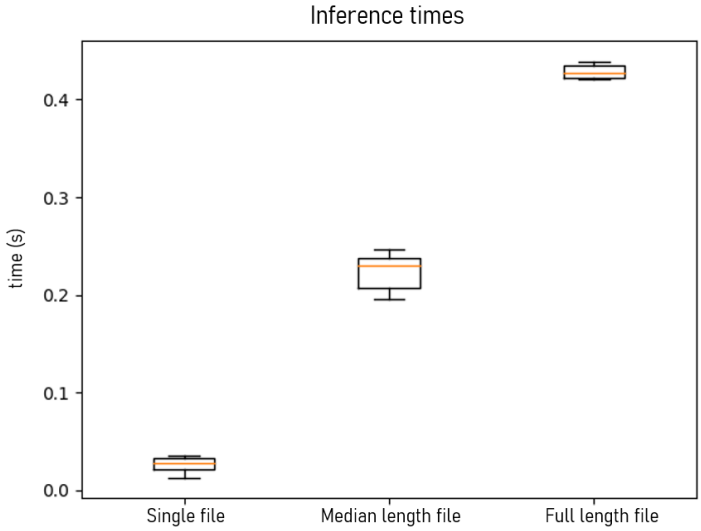

Figure 6. Algorithm's inference times (in seconds) for a single file classification, silence removal, and classification of a varying size of 8–14 s audio file (median length file) and 30 s full-length-file, respectively.

The algorithm was trained and evaluated on an Intel Core i5-7600 @ 3.5GHz processor.

4. Discussion

As has been mentioned in the introduction, many related-content works have been published in the last years aiming to detect and classify sound events from the use of an inhaler. Unfortunately, a one-to-one comparison between other publications cannot be approached. First, the data used for the algorithms' evaluation were not the same. Within this work's scopes, a new dataset was created since data from previous researches were not publicly available. In addition, different devices were used for the data capture, with different sampling rates and depth. Finally, the procedure itself is not comparable because others extracted a different set of features, used different representations of the signal, and aimed at solving different classification tasks (different number of targeted classes).

Nevertheless, the model can be generalized to other commonly used DPIs. For example, for an assessment of the Diskus inhaler usage, the evaluation of the procedure would be completed if the algorithms can detect two *Button press* ("clicks") in a row, *Exhale*, an *Inhale* event (but not *Actuation* since there is no such evaluation mechanism in this inhaler), again *Exhale*, and finally another *Button press*. For other inhalers, such as Turbuhaler, Respimat, Nexthaler, the mechanisms of the proposed method could also be applied with differentiation in the sequence of the events, but with no readjustments of the model's audio-based classification method.

A limitation of this study could be the fact that the assessment is performed using only audio data, but under some circumstances, data from a camera or an accelerometer could be used to cover all the possible cases. If a patient misplaces the device in their mouth, for example, with the button facing down and not up the audio analysis method could not locate this error, while camera and accelerometer could. Another potential limitation is that in this work no extreme scenarios regarding

background noises where investigated for the data capture, e.g., noises existing in a city with traffic. Therefore, there should be a more extensive data collection process.

5. Conclusions

In this work, a system architecture designed for efficient classification of audio signals into four major categories, i.e., *Actuation*, *Inhale*, *Button press*, and *Exhale*, is proposed. The algorithm focuses in the usage of traditional classifiers and not deep neural networks, focusing on the manually engineered features. The segmentation of the signal is performed not with the use of a sliding window but with the implementation of silence removal as proposed in [30]. We demonstrated that our network achieves high performance, while being able to generalize well, since a big amount of patients contributed to the training and testing process of the algorithm. A post-processing step of peak detection is applied to further improve the prediction from the automatic classification procedure, giving an improvement of 5.58% in the F1-score. In addition, the network is less demanding in terms of computing resources needed for training and is introducing a very little processing delay, achieving near real-time inference.

6. Future Work

The proposed framework takes as input a raw acoustic signal and can provide information regarding the sequence of inhale events and detect possible patient technique errors. As a future step, a smartphone application will be developed in order to provide real-time feedback to the end user. The patient will record the inhalation procedure through the mobiles microphone; these audio data will be transferred to a server where the silence removal and classification method will be applied. Finally, in less than half a second (as shown in Figure 6 for a 30 s long file) the evaluation of the followed procedure with be displayed to the patient's phone. Information regarding the missed steps could be displayed (in addition to the produced figure), as plain text or as interactive video instructions. In this way, the inhaler usage experience will be enhanced through intuitive interfaces and patients could assess their inhaler technique in an engaging manner.

In addition, this mobile app, which can communicate with a clinical platform, will be a helpful tool for supervising doctors. The results from the procedure assessment can be reported to the doctors at the next visit to their office, where all the performed actions can be logged in a history file. This can be done via a remote monitoring platform that displays all the information of each patient. Doctors can have real-time access, following the patient's progress, while being able to provide relevant feedback and personalized guidance.

A mobile monitoring system with the proposed methodology included can play an important role in self-management of the disease and in the prevention of exacerbation symptoms that in many cases can be fatal. In addition, it can result in better quality of services from the healthcare system, continuous monitoring, and better management of the diagnosis from the supervising doctors.

Author Contributions: Conceptualization, A.-C.E. and A.V.; methodology, A.-C.E. and A.V.; software, A.-C.E. and A.V.; validation, A.-C.E., A.V., and A.L.; formal analysis, A.-C.E. and A.V.; investigation, A.-C.E., A.V., and A.L.; resources, A.-C.E. and A.V.; data curation, A.-C.E.; writing—original draft preparation, A.-C.E.; writing—review and editing, A.-C.E. and A.V.; visualization, A.-C.E. and A.V.; supervision, A.V., A.L., K.V., and D.T.; project administration, K.V. and D.T.; funding acquisition, K.V. and D.T. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE under grant agreement No. T1EDK-03832 (Take-A-Breath project).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Pocket Guide for Asthma Management and Prevention. Available online: https://ginasthma.org/wp-content/uploads/2019/04/GINA-2019-main-Pocket-Guide-wms.pdf (accessed on 10 September 2020).
2. Vos, T.; Abajobir, A.A.; Abate, K.H.; Abbafati, C.; Abbas, K.M.; Abd-Allah, F.; Abdulkader, R.S.; Abdulle, A.M.; Abebo, T.A.; Abera, S.F.; et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **2017**, *390*, 1211–1259.
3. Cukic, V.; Lovre, V.; Dragisic, D.; Ustamujic, A. Asthma and chronic obstructive pulmonary disease (COPD)–differences and similarities. *Mater. Socio Medica* **2012**, *24*, 100.
4. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. Available online: https:https://goldcopd.org/wp-content/uploads/2019/11/GOLD-2020-REPORT-ver1.0wms.pdf (accessed on 10 September 2020).
5. Masoli, M.; Fabian, D.; Holt, S.; Beasley, R.; Global Initiative for Asthma (GINA) Program. The global burden of asthma: Executive summary of the GINA Dissemination Committee report. *Allergy* **2004**, *59*, 469–478. [PubMed]
6. The Inhaler Error Steering Committee; Price, D.; Bosnic-Anticevich, S.; Briggs, A.; Chrystyn, H.; Rand, C.; Scheuch, G.; Bousquet, J. Inhaler competence in asthma: Common errors, barriers to use and recommended solutions. *Respir. Med.* **2013**, *107*, 37–46.
7. Ocakli, B.; Ozmen, I.; Tunçay, E.A.; Gungor, S.; Altinoz, H.; Adiguzel, N.; Sak, Z.A.; Gungor, G.; Karakurt, Z.; Arbak, P. A comparative analysis of errors in inhaler technique among COPD versus asthma patients. *Int. J. Chronic Obstr. Pulm. Dis.* **2018**, *13*, 2941.
8. Lindh, A.; Theander, K.; Arne, M.; Lisspers, K.; Lundh, L.; Sandelowsky, H.; Ställberg, B.; Westerdahl, E.; Zakrisson, A.B. Errors in inhaler use related to devices and to inhalation technique among patients with chronic obstructive pulmonary disease in primary health care. *Nurs. Open* **2019**, *6*, 1519–1527.
9. Haughney, J.; Price, D.; Barnes, N.C.; Virchow, J.C.; Roche, N.; Chrystyn, H. Choosing inhaler devices for people with asthma: Current knowledge and outstanding research needs. *Respir. Med. CME* **2010**, *3*, 125–131. [CrossRef]
10. Fink, J.B.; Rubin, B.K. Problems with inhaler use: A call for improved clinician and patient education. *Respir. Care* **2005**, *50*, 1360–1375.
11. Lewis, A.; Torvinen, S.; Dekhuijzen, P.; Chrystyn, H.; Watson, A.; Blackney, M.; Plich, A. The economic burden of asthma and chronic obstructive pulmonary disease and the impact of poor inhalation technique with commonly prescribed dry powder inhalers in three European countries. *BMC Health Serv. Res.* **2016**, *16*, 251.
12. Batterink, J.; Dahri, K.; Aulakh, A.; Rempel, C. Evaluation of the use of inhaled medications by hospital inpatients with chronic obstructive pulmonary disease. *Can. J. Hosp. Pharm.* **2012**, *65*, 111.
13. Pritchard, J.N.; Nicholls, C. Emerging technologies for electronic monitoring of adherence, inhaler competence, and true adherence. *J. Aerosol Med. Pulm. Drug Deliv.* **2015**, *28*, 69–81. [CrossRef]
14. Sulaiman, I.; Seheult, J.; Sadasivuni, N.; MacHale, E.; Killane, I.; Giannoutsos, S.; Cushen, B.; Mokoka, M.C.; Bhreathnach, A.S.; Boland, F.; et al. The impact of common inhaler errors on drug delivery: Investigating critical errors with a dry powder inhaler. *J. Aerosol Med. Pulm. Drug Deliv.* **2017**, *30*, 247–255. [CrossRef]
15. Azouz, W.; Campbell, J.; Stephenson, J.; Saralaya, D.; Chrystyn, H. Improved metered dose inhaler technique when a coordination cap is used. *J. Aerosol Med. Pulm. Drug Deliv.* **2014**, *27*, 193–199. [CrossRef]
16. Akinbami, L.J.; Simon, A.E.; Rossen, L.M. Changing trends in asthma prevalence among children. *Pediatrics* **2016**, *137*, e20152354. [CrossRef]
17. Krishnan, V.; Diette, G.B.; Rand, C.S.; Bilderback, A.L.; Merriman, B.; Hansel, N.N.; Krishnan, J.A. Mortality in patients hospitalized for asthma exacerbations in the United States. *Am. J. Respir. Crit. Care Med.* **2006**, *174*, 633–638. [CrossRef]
18. AAAA&I. Dry Powder Inhalers. Available online: https://www.aaaai.org/conditions-and-treatments/conditions-dictionary/dry-powder-inhalers (accessed on 29 July 2020).

19. Javadzadeh, Y.; Yaqoubi, S. Therapeutic nanostructures for pulmonary drug delivery. In *Nanostructures for Drug Delivery*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 619–638.
20. Genuair, B. INN-Aclidinium Bromide. Available online: https://www.ema.europa.eu/en/documents/product-information/bretaris-genuair-epar-product-information_en.pdf (accessed on 29 July 2020).
21. Berkenfeld, K.; Lamprecht, A.; McConville, J.T. Devices for dry powder drug delivery to the lung. *Aaps Pharmscitech* **2015**, *16*, 479–490. [CrossRef]
22. Dal Negro, R.W.; Turco, P.; Povero, M. Patients' usability of seven most used dry-powder inhalers in COPD. *Multidiscip. Respir. Med.* **2019**, *14*, 30. [CrossRef]
23. Taylor, T.E.; Zigel, Y.; De Looze, C.; Sulaiman, I.; Costello, R.W.; Reilly, R.B. Advances in audio-based systems to monitor patient adherence and inhaler drug delivery. *Chest* **2018**, *153*, 710–722. [CrossRef]
24. Holmes, M.S.; D'arcy, S.; Costello, R.W.; Reilly, R.B. Acoustic analysis of inhaler sounds from community-dwelling asthmatic patients for automatic assessment of adherence. *IEEE J. Transl. Eng. Health Med.* **2014**, *2*, 1–10. [CrossRef]
25. Taylor, T.E.; Holmes, M.S.; Sulaiman, I.; D'Arcy, S.; Costello, R.W.; Reilly, R.B. An acoustic method to automatically detect pressurized metered dose inhaler actuations. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 4611–4614.
26. Kikidis, D.; Votis, K.; Tzovaras, D. Utilizing convolution neural networks for the acoustic detection of inhaler actuations. In Proceedings of the 2015 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 19–21 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–4.
27. Nousias, S.; Lakoumentas, J.; Lalos, A.; Kikidis, D.; Moustakas, K.; Votis, K.; Tzovaras, D. Monitoring asthma medication adherence through content based audio classification. In Proceedings of the 2016 IEEE symposium series on computational intelligence (SSCI), Athens, Greece, 6–9 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–5.
28. Lalos, A.S.; Lakoumentas, J.; Dimas, A.; Moustakas, K. Energy efficient monitoring of metered dose inhaler usage. *J. Med Syst.* **2016**, *40*, 285. [CrossRef]
29. Butterworth, S. On the theory of filter amplifiers. *Wirel. Eng.* **1930**, *7*, 536–541.
30. Giannakopoulos, T. pyaudioanalysis: An open-source python library for audio signal analysis. *PLoS ONE* **2015**, *10*, e0144610. [CrossRef]
31. Zheng, F.; Zhang, G.; Song, Z. Comparison of different implementations of MFCC. *J. Comput. Sci. Technol.* **2001**, *16*, 582–589. [CrossRef]
32. Chen, C.H. *Signal Processing Handbook*; CRC Press: Boca Raton, FL, USA, 1988; Volume 51.
33. Sakhnov, K.; Verteletskaya, E.; Simak, B. Dynamical energy-based speech/silence detector for speech enhancement applications. In Proceedings of the World Congress on Engineering, Citeseer, London, UK, 1–3 July 2009; Volume 1, p. 2.
34. Dubnov, S. Generalization of spectral flatness measure for non-gaussian linear processes. *IEEE Signal Process. Lett.* **2004**, *11*, 698–701. [CrossRef]
35. Peeters, G. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Proj. Rep.* **2004**, *54*, 1–25.
36. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
37. Tzanetakis, G.; Cook, P. Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.* **2002**, *10*, 293–302. [CrossRef]
38. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
39. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
40. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
41. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]