Article

Cough Detection Using Acceleration Signals and Deep Learning Techniques

Daniel Sanchez-Morillo [1,2,*] , Diego Sales-Lerida [1,2] , Blanca Priego-Torres [1,2] and Antonio León-Jiménez [2,3]

1   Department of Automation Engineering, Electronics and Computer Architecture and Networks, School of Engineering, University of Cadiz, 11519 Cádiz, Spain; diego.lerida@uca.es (D.S.-L.); blanca.priego@uca.es (B.P.-T.)
2   Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), 11009 Cádiz, Spain
3   Pulmonology Department, Puerta del Mar University Hospital, 11009 Cádiz, Spain
*   Correspondence: daniel.morillo@uca.es

Abstract: Cough is a frequent symptom in many common respiratory diseases and is considered a predictor of early exacerbation or even disease progression. Continuous cough monitoring offers valuable insights into treatment effectiveness, aiding healthcare providers in timely intervention to prevent exacerbations and hospitalizations. Objective cough monitoring methods have emerged as superior alternatives to subjective methods like questionnaires. In recent years, cough has been monitored using wearable devices equipped with microphones. However, the discrimination of cough sounds from background noise has been shown a particular challenge. This study aimed to demonstrate the effectiveness of single-axis acceleration signals combined with state-of-the-art deep learning (DL) algorithms to distinguish intentional coughing from sounds like speech, laugh, or throat noises. Various DL methods (recurrent, convolutional, and deep convolutional neural networks) combined with one- and two-dimensional time and time–frequency representations, such as the signal envelope, kurtogram, wavelet scalogram, mel, Bark, and the equivalent rectangular bandwidth spectrum (ERB) spectrograms, were employed to identify the most effective approach. The optimal strategy, which involved the SqueezeNet model in conjunction with wavelet scalograms, yielded an accuracy and precision of 92.21% and 95.59%, respectively. The proposed method demonstrated its potential for cough monitoring. Future research will focus on validating the system in spontaneous coughing of subjects with respiratory diseases under natural ambulatory conditions.

Keywords: cough; deep learning; accelerometer; mechano-acoustic signals; respiratory sounds

1. Introduction

Cough has been the subject of study in several disciplines of medicine and is considered a common healthcare problem with an associated high socioeconomic burden [1]. Chronic cough affects approximately 10% of the adult population and can severely impair quality of life [2]. Cough is the most common symptom that causes people to seek medical attention [3]. However, cough is considered not only a symptom but also a predictive indicator of acute adverse health events [4]. It is connected with many common respiratory diseases, such as lung cancer, tuberculosis, chronic pulmonary obstructive disease (COPD), asthma, and infections affecting the lower respiratory tract, including coronavirus disease 2019 (COVID-19); and it is associated with exacerbations, impaired lung function, and risk of death [5,6]. In addition, cough has been associated with psychological morbidity [7], and its assessment is also important for evaluating the therapy response [8].

Although quantifying cough frequency is useful as a marker of cough severity and would allow the assessment of cough as a trigger for adverse health events, cough frequency monitoring has not been incorporated into routine clinical practice and continues to be a research instrument [9]. Among the barriers indicated in the recent literature, it has been

suggested that cough telemonitoring systems need to be refined and validated for optimal clinical utility [10].

The measurement and monitoring of cough have been conventionally carried out using questionnaires, where the patients self-report the frequency and severity of their cough episodes. This is the case of the Leicester cough questionnaire (LCQ), the cough symptom score (CSS), and the cough-specific quality of life questionnaire (CQLW) [11]. However, the collection of signs and symptoms using questionnaires is subjective since it is biased by patient perception [12]. Available objective methods for assessing cough have been limited to cough sensitivity assessments, which require chemical or mechanical stimulation [13].

Since the automatic detection and characterization of cough episodes has been referred to as a strategy of great interest that can improve disease monitoring [14], trends in respiratory health monitoring have shifted towards more non-invasive and automatic approaches. In the last decade, the audio signal has been used for cough characterization, using 24 h monitoring systems to count episodes (e.g., in studies of drugs to treat cough). One or more body-worn or lapel-type microphones, and sensors embedded in mobile phones have been and continue to be used for this purpose [15,16]. The audio signal recorded by the microphones has been processed by trained algorithms to reduce noise, discard periods of silence, and classify cough events with a high probability of success [17].

The discrimination of cough sounds from other noise (e.g., throat clearing) and speech signals has been shown to be a particular challenge. In addition, despite its long-standing use, commercially available solutions generally still rely on a manual component [14]. To overcome this limitation, sound collection using smartphones and machine learning algorithms has shown promise in effectively measuring cough [18].

As an alternative to microphones, other novel sensors like chest-laminated electronic skin (e-skin) have been proposed for reliable cough detection [19]. In this line, automatic cough detection methods using accelerometers are being explored to enable continuous monitoring of cough frequency and severity. Technologies like accelerometer-based bed occupancy detection systems [20,21] or triaxial accelerometers combined with stretchable strain sensors [22] have shown promising results in accurately detecting cough episodes for long-term monitoring [23]. Additionally, advancements in artificial intelligence have demonstrated promising potential for improving the accuracy and efficiency of cough detection systems [24]. However, the performance of these models in realistic environments remains limited by their significant false positive rate.

This study aimed to show how an accelerometer-based system combined with state-of-the-art deep learning techniques can enable the automatic detection of cough episodes. Acceleration body-worn sensors can more effectively discriminate coughs from external sounds, and the application of artificial intelligence techniques can contribute to the mitigation of false positives. Deep learning approaches were thoroughly explored and compared using the acceleration signals (vibrations) collected using an accelerometer placed on the suprasternal notch.

2. Materials and Methods

### 2.1. Participants and Data Acquisition Protocol

To train and validate the artificial intelligence-based models, experimental data were collected from a group of 23 participants (14 men and 9 women). The main objective was the automatic detection of cough, differentiating it from other acoustic events such as speech, laugh, throat clearing, or background noise.

All subjects gave informed consent for inclusion before participating in the study. The study was conducted by the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Biomedical Research of Cádiz (Project identification code CICERONE, 29.23).

The acceleration signals of each participant were recorded in a single file using a common data structure that facilitated the subsequent separation and labeling of each

signal segment. Each recording was carried out with the subject in a seated position and included 10 voluntary coughing episodes, 5 s of speech, two seconds of laughing, two seconds of throat clearing, and two pauses of silence in which the subject's breathing was recorded. Groups of different sounds were separated by two seconds of silence. A user interface was designed to guide the participants during the signal acquisition.

Figure 1 shows an example of the accelerometer signal recorded during an experimental session, including cough and non-cough episodes (silence, speech, laugh, and throat clearing).
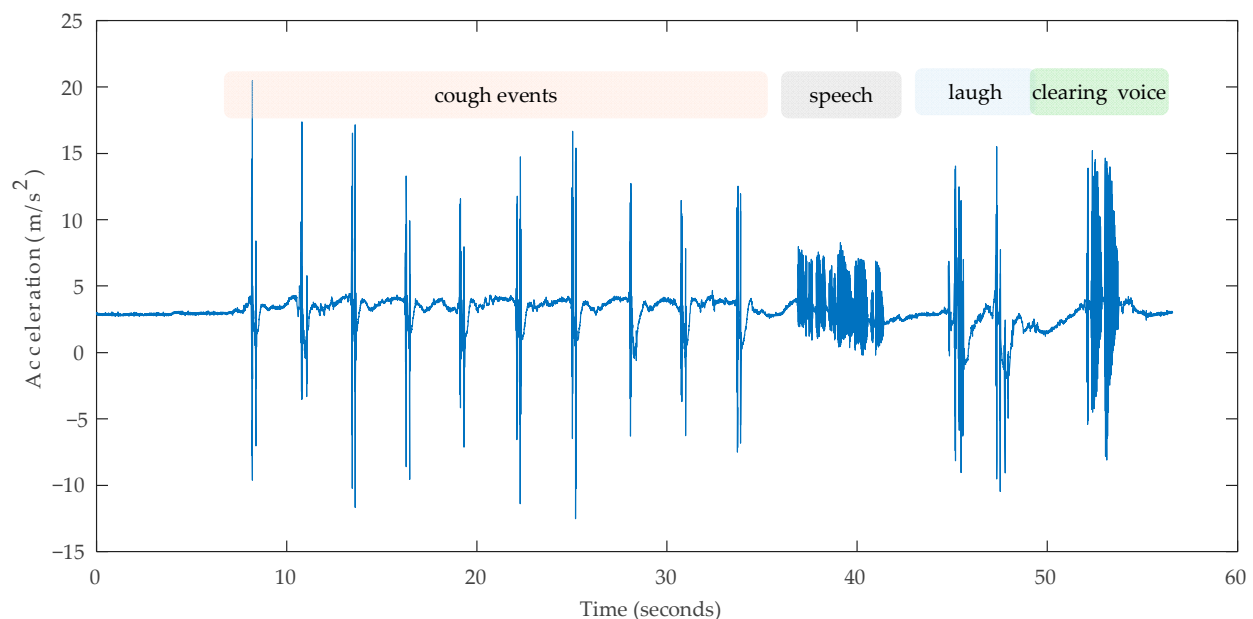
Figure 1. An example of the accelerometer signal recorded in one of the experimental sessions. The different types of signals recorded over time in each experiment (cough, speech, laugh, and clearing voice noise) are labeled.

2.2. Hardware

The data acquisition system comprised an accelerometer sensor, with a housing for optimal positioning, and a microprocessor system (Raspberry Pi Zero W. The Raspberry Pi Zero W unit (Broadcom BCM2835 processor with 1 GHz ARM11 core, 512 MB of RAM, and Linux operating system) was selected as the microprocessor due to its low power consumption, wireless connectivity, and small size. An SD card allowed the storage of the recorded signals. The system was powered by a rechargeable LiPo battery, connected to the microprocessor through a DC/DC converter charger. The assembly was installed in a 3D-printed plastic enclosure designed specifically for this purpose. The acceleration signals were acquired using the low-power consumption MMA8452Q 3-axis acceleration sensor (Freescale Semiconductor, Austin, TX, USA). The MMA8452Q sensor has a resolution of 12 bits and supports a low power consumption mode that extends battery life. It can measure accelerations in the range $[-2g, +2g]$, offers I2C communications, and supports a maximum sampling rate of 800 Hz. It is often used for medical monitoring and rehabilitation by providing real-time movement data. Only the *z*-axis was used in this study to acquire acceleration signals perpendicular to the skin surface at the point of contact. The optimal location for the sensor placement was estimated to be around the thyroid cartilage at the front of the neck, according to a previous study by the authors [25]. The sensor was attached to the participant's skin using an adhesive electrode. The connection between the sensor and the microprocessor unit was made using the I2C communications protocol. Figure 2 illustrates the prototype of the data acquisition system implemented in this study.
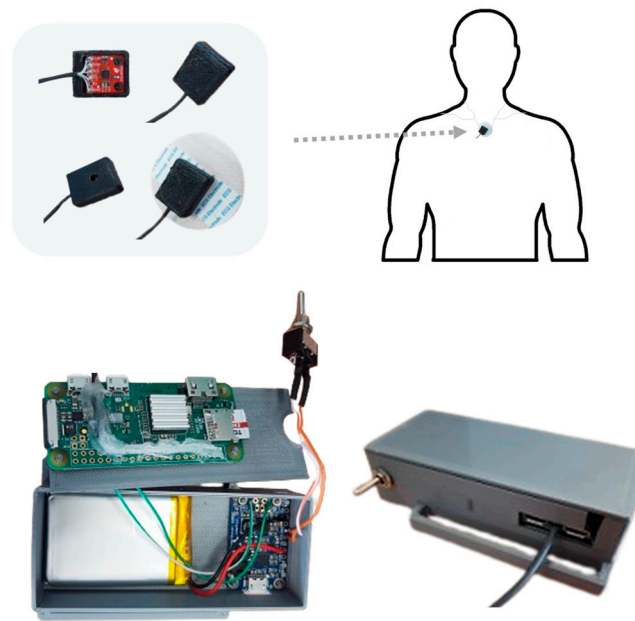
Figure 2. Detail of the prototype for signal acquisition.

2.3. Signal Processing and Data Augmentation

The acceleration signals were acquired using a sampling rate of 800 Hz. The reading was carried out by the microprocessor using interruptions. The acquisition and recording process was controlled using the Python 3 programming language. The raw signal was filtered to eliminate unwanted movements of the subject and remove the noise due to the sensor rubbing against the skin. A high-pass FIR filter of order 10 with a cutoff frequency of 50 Hz was used.

The number of 1 s intervals per hour that contain a cough has been proposed as a cough monitoring index [26]. Accordingly, the cough and non-cough samples were segmented manually into 1 s segments, leaving, if possible, silent intervals before and after. No overlapping was applied. Cough records of less than 1 s were completed using the zero-padding technique.

Random sequential augmentations were applied to increase the number of samples of cough events. Pitch shifting, time-scale modification, time shifting, noise addition, and amplitude control were used to increase the number of cough and non-cough samples by a factor of 10 and 2, respectively.

Table 1 shows the distribution of participants, gender, and samples in the original and augmented sets.

Table 1. Distribution of participants, recordings, and samples in the original and augmented sets.

	Original Dataset			**Augmented Dataset**		
Participants			23			
Female (%)			39.1%			
	Cough	Non-Cough	Total	Cough	Non-Cough	Total
Samples	206	1011	1217	2266	2022	4288
Duration	3 min 50 s	16 min 52 s	20 min 42 s	37 min 46 s	33 min 42 s	71 min 28 s

2.4. Time–Frequency Representations and Feature Extraction

2.4.1. Two-Dimensional Spectrogram Feature Extraction

To train the 2D convolutional neural networks (CNN), the VGGish model, and the deep convolutional neural networks (DCNNs), we used two-dimensional (2D) representations, common in audio signal processing, to capture and describe the relevant features of the

acceleration signals. Representations such as mel spectrograms [27], Bark and equivalent rectangular bandwidth (ERB) spectra [28], wavelet scalograms [29], and kurtograms [30] were considered because they provide detailed insight into the temporal and frequency structure of signals.

- Mel, ERB, and Bark spectrograms

Comparing the mel, ERB, and Bark spectrograms applied to the accelerometer signal can provide a deeper understanding of how different frequency scales can influence the identification of cough patterns. This can lead to more precise and efficient detection, improving health monitoring applications.

Log-mel spectrograms are representations of the signal spectral energy, where the energy at each frequency is calculated at discrete time intervals. The mel scale is applied to the frequencies before taking the logarithm of the energy. The ERB spectrum also represents the signal spectral energy. In this case, the energy at each frequency is grouped into bands of equivalent rectangular width. The Bark spectrum provides the signal spectral energy using Bark bands.

A 25 ms Hanning window with 10 ms overlapping was used to calculate the mel, Bark, and ERB spectrograms. The FFT length was 512 and the number of bands was set to 50.

- Kurtogram, scalogram, and signal envelope

The kurtogram captures the kurtosis of a signal at different time and frequency scales. Kurtosis is a measure of the shape of the signal distribution and can provide useful information about the temporal and spectral characteristics of the signal.

The wavelet scalogram represents the signal as a function of time and frequency, using wavelets to analyze the signal characteristics at different time and frequency scales. The analytic Morse wavelet and 12 voices per octave were used in this study to calculate the continuous wavelet transform.

Finally, the upper and lower envelopes of the acceleration signal were used. The envelopes were determined using spline interpolation over local maxima separated by at least 10 samples.

2.4.2. One-Dimensional Spectrum Feature Extraction

Some of the artificial intelligence techniques used in this study required one-dimensional (1D) signal features as inputs. Mel Frequency Cepstral Coefficients (MFCCs), delta, and delta2 MFCCs were calculated for this purpose. These coefficients capture the signal spectral characteristics by mapping the energy of the frequency spectrum into a nonlinear mel scale. MFCCs capture information about the spectral shape and content of the signal, making them valuable for classification tasks. Delta MFCCs represent the rate of change of adjacent MFCCs over time and the delta2 MFCC is the differential of the MFCC delta coefficients (acceleration). These parameters provide information useful for distinguishing between signals with different spectral change patterns over time [31]. In total, 13 MFCC, 13 delta MFCC, and 13 delta2 MFCC parameters were calculated.

In addition, spectral (spectral centroid, spectral roll-off, spectral slope, spectral spread, harmonic ratio) and time-domain features (zero-crossing rate and short-time energy) were estimated [32]. To calculate the spectral parameters, a Hamming window with 50% overlapping was used.

The spectral centroid (C), roll-off, slope (m), and spread (σ) were calculated on the one-sided mel spectrum. The spectral centroid and the spectral spread are statistical spectrum moments and represent the spectral gravity center and the spread of the spectrum around the mean value, respectively. The spectral slope captures the slope of the signal spectrum and reflects the energy distribution between low and high frequencies. The spectral roll-off indicates the frequency f_K below which 95% of the total spectral energy is contained.

These parameters were calculated according to Equations (1)–(4):

$$C = \frac{\sum_{n=1}^{N} f_i S_i}{\sum_{n=1}^{N} S_i} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{n=1}^{N} (f_i - C)^2 S_i}{\sum_{n=1}^{N} S_i}} \tag{2}$$

$$m = \frac{N \sum_{n=1}^{N} (f_i \, S_i) - \sum_{n=1}^{N} f_i \, \sum_{n=1}^{N} S_i}{N \sum_{n=1}^{N} f_i{}^2 - \left(\sum_{n=1}^{N} f_i\right)^2} \tag{3}$$

$$\sum_{n=1}^{k} S_i \geq 0.95 \sum_{n=1}^{N} S_i \tag{4}$$

where f_i are the mel scale frequencies, S_i is the amplitude of the spectrum at the frequency f_i, N is the number of points in the spectrum, and k is the spectral roll-off point (f_K).

The harmonic ratio (R_H) quantifies the proportion of harmonic (periodic) content to the total energy in the signal and was calculated using the normalized signal autocorrelation according to [33].

The short-time energy was computed in the time domain according to Equation (5):

$$sTE = \sum_{n=0}^{N-1} x_w(n)^2 \tag{5}$$

where $x_w(n)$ is the windowed signal.

The zero-crossing rate was estimated as the rate at which a signal changed from positive to zero to negative or from negative to zero to positive.

To assess the effect of the window size, the performance of models for window sizes that ranged from 25 ms to 400 ms was explored. In total, 46 1D-features were calculated for each sample. Features were standardized.

2.5. Deep Learning Models

In the last decade, there has been a growing interest in applying artificial intelligence techniques for respiratory sound classification tasks such as cough detection and respiratory disorder identification [34–37]. In general, deep learning models have demonstrated great potential to achieve an accuracy and performance higher than those of traditional machine learning models.

In this study, four categories of deep learning network models were used as binary classifiers capable of detecting cough events: recurrent neural networks (RNNs), convolutional neural networks (CNNs), the VGGish network, and deep convolutional neural networks (DCNNs). Figure 3 shows the time–frequency features used for each DL family.

2.5.1. Recurrent Neural Networks (RNNs)

RNNs, despite their sequential nature, can be deep if they have multiple layers. Deep RNNs have multiple recurrent layers stacked on top of each other. RNNs process input data step-by-step and have loops within their structure, allowing information to persist over time. This makes these models suitable for tasks such as time series prediction, natural language processing, and voice disorder detection [38].

The Long Short-Term Memory (LSTM) model was used in this study. The LSTM architecture has been used in the context of cough sounds, tuberculosis, and COVID-19 with promising results [39,40]. An LSTM model is a type of RNN designed to handle the vanishing gradient problem. It is capable of learning long-term dependencies in sequential data, making it suitable for processing sequences of cough sounds over time.

The LSTM model used in this study was designed with an LSTM layer and two fully connected layers (Figure 4). The input layer is fed with the 46 spectral and time domain features extracted from each 1 s segment of the acceleration signal, as described

in Section 2.4.2. The second fully connected layer feeds into a SoftMax classifier with two class labels.
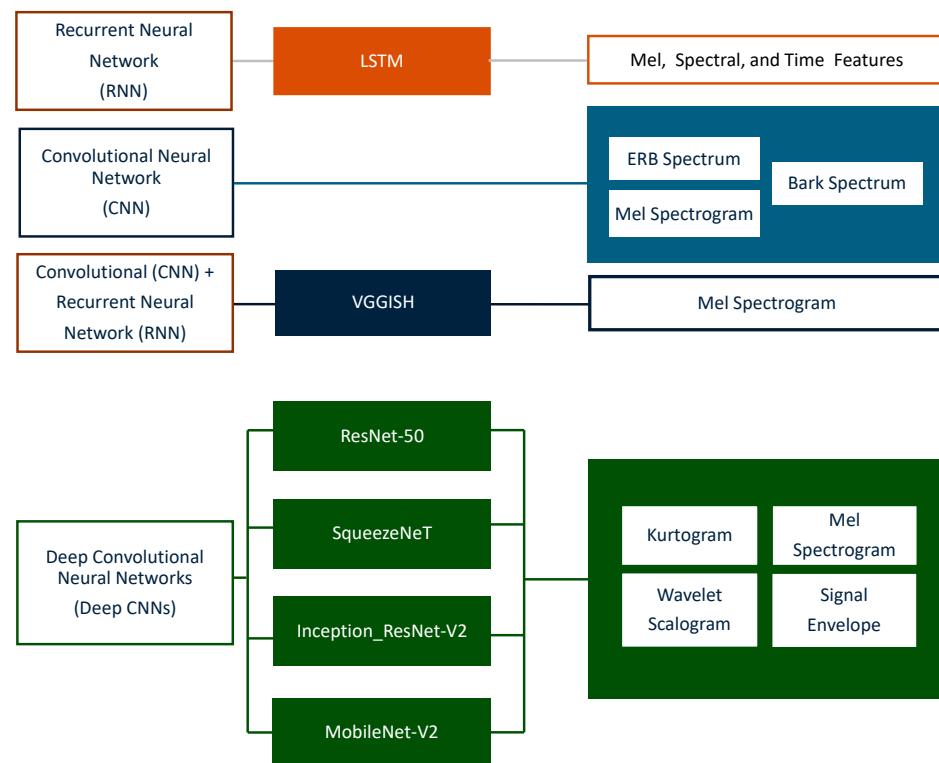
Figure 3. Deep learning models, features, and time–frequency representations used in the study.
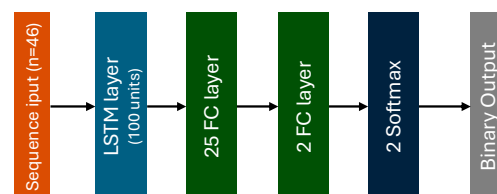
Figure 4. LSTM model architecture. FC: Fully connected.

2.5.2. Convolutional Neural Networks (CNNs)

A CNN is a type of feed-forward neural network with convolutional layers followed by pooling layers to learn local features in the input data. CNNs have a relatively straightforward architecture and are primarily used for image recognition and classification tasks. However, they have been adapted for processing other data types, including cough sounds [40,41].

Generally, a CNN model contains two main parts: a convolution layer and a fully connected layer. The convolution layer performs feature extraction. In contrast, the fully connected layer utilizes the output of the convolution layer to generate a compelling image description.

In this study, the CNN model detailed in Figure 5 was used for the cough classification task. This architecture has been proposed recently for heart sound classification tasks [42]. It includes five convolutional layers and one fully connected layer. The first three convolutional layers are followed by overlapping max-pooling layers. The fifth convolutional layer is linked to the fully connected layer, which feeds into a SoftMax classifier with two class labels. ERB, MEL, and Bark spectrograms were used as input images (98×50) to this CNN model.
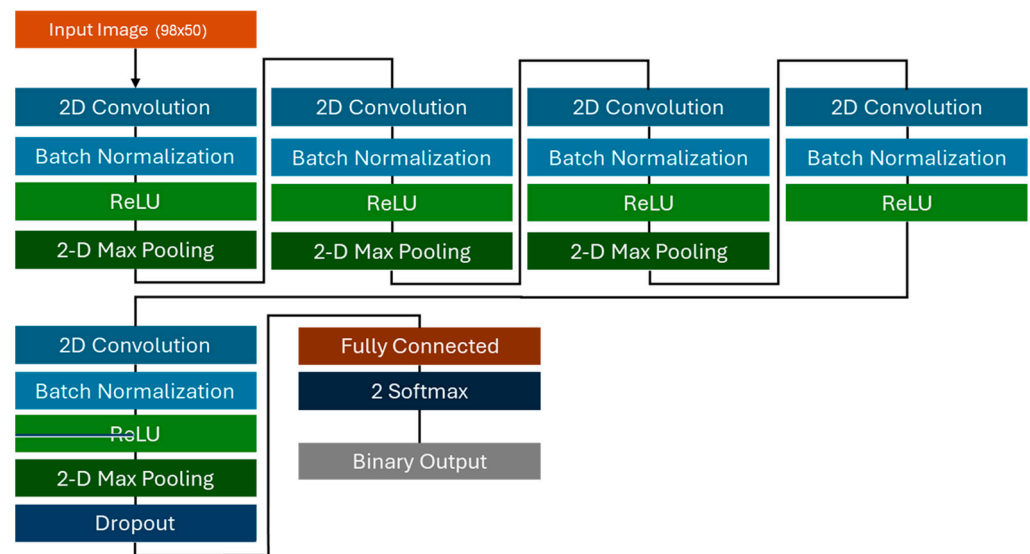
Figure 5. Architecture of the CNN model.

2.5.3. Mixed RNN-CNN Approach: VGGish

VGGish is a deep learning model that utilizes both convolutional and recurrent neural network architectures. It was developed by Google for audio feature extraction and is based on the VGG (Visual Geometry Group) architecture. The VGGish model operates on log-mel spectrograms and uses 2D convolutional, pooling, and fully connected layers to extract relevant features from signal spectrograms (Figure 6). VGGIsh has been used very recently as a feature extractor for COVID-19 cough classification [43]. The pre-trained VGGish model was used and fine-tuned by letting it iterate on our dataset.
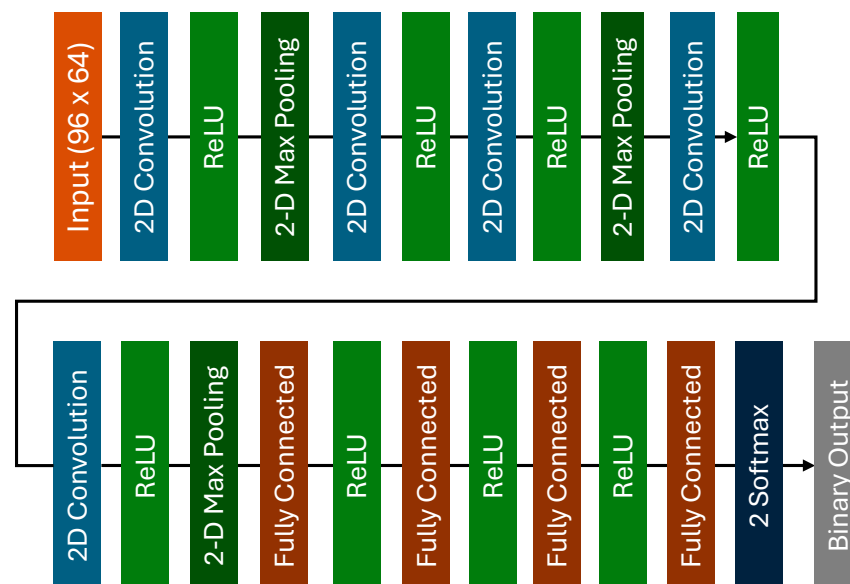
Figure 6. The architecture of the VGGish model.

Before feeding the VGGish model, the mechano-acoustic signals were resampled to 16,000 Hz, and an array of mel spectrograms was computed to produce a fixed-length feature vector (96×64) for each signal segment. The window size was 25 ms and the overlap percentage between consecutive mel spectrograms was 50%. These feature vectors were used as inputs to tune a pre-trained VGGish classifier.

2.5.4. Deep Convolutional Neural Networks (DCNNs)

DCNNs refer specifically to CNN architectures with a significant number of layers, which are capable of learning highly abstract and complex features from raw data. DCNNs typically consist of many convolutional layers, sometimes interspersed with pooling layers, and followed by fully connected layers. In this study, we trained and validated four DCNN networks, namely ResNet-50, SqueezeNet, Inception-Resnet-V2, and MobileNet-V2.

ResNet-50 is a DCNN that uses residual connections for extracting complex features from input data. SqueezeNet was designed to be lightweight and computationally efficient, using 1×1 convolutions to reduce the number of parameters, and enabling efficient implementation on resource-constrained devices. Inception-Resnet-V2 architecture combines the idea of Inception modules with residual connections, allowing for more efficient learning of complex features at different scales. Finally, the MobileNet-V2 architecture utilizes depth-wise separable convolutions to achieve a balance between accuracy and efficiency. These networks have been very recently used for tuberculosis and COVID-19 detection using cough sounds [40,44,45]. We take the pre-trained networks and use them as a starting point to learn the task of cough detection using transfer learning.

2.6. Validation

Leave-One-Subject-Out Cross-Validation (LOSOCV) was applied to evaluate the performance of deep learning models. LOSOCV is commonly used in contexts involving subject-specific data and is based on dividing the dataset by subjects, ensuring that all data from one subject is used as the test set, while data from the remaining subjects are used as the training set. This process was repeated for each subject in the dataset, ensuring that each subject's data were used once for testing. The average model performance across all iterations was then calculated to obtain the overall metrics.

All deep learning models were trained and validated using MATLAB R2023a and a 12th Gen Intel(R) Core (TM) i7-12700K 3.60 GHz, equipped with an NVIDIA GeForce RTX 3070Ti GPU. Some tests were carried out to evaluate the hyperparameters that led to stable training processes and high-performance metrics. The learning rate and the batch size were set to 1×10^{-4} and 64, respectively. For each training run, the model version with the best validation loss was stored and evaluated.

2.7. Performance Metrics

To ensure consistency and generalization of each resulting model, metrics widely used to measure the performance in machine learning classifiers were calculated. These metrics included accuracy (Acc), precision (Pr), sensitivity (Se), specificity (Sp), false positive rate (FPR), and the Cohen's Kappa index. Acc, Pr, Se, Sp, and FPR were calculated using the confusion matrix (Equations (6)–(10)). The Cohen's Kappa index was estimated using Equations (11) and (12). Table 2 shows the confusion matrix for a binary classification problem.

Table 2. Confusion matrix.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

True positive (TP) and true negative (TN) represent the number of samples in the positive class and negative class that are correctly classified, respectively. False negative (FN) and false positive (FP) represent the samples that are incorrectly classified in the positive and negative class.

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Pr = \frac{TP}{TP + FP} \tag{7}$$

$$Se = \frac{TP}{TP + FN} \tag{8}$$

$$Sp = \frac{TN}{TN + FP} \tag{9}$$

$$FPR = \frac{FP}{FP + TN} \tag{10}$$

$$Kappa = \frac{Acc - Pe}{1 - Pe} \tag{11}$$

$$Pe = \frac{(TP + FN) * (TP + FP) + (FP + TN) * (FN + TN)}{(TP + FP + FN + TN)^2} \tag{12}$$

Ideally, the FPR should be low to prevent a significant number of non-cough episodes from being detected as coughs. The closer the rest of the metrics are to 100%, or 1 in the case of the Kappa index, the better the model performance.

All metrics were determined with 95% confidence intervals (CIs). The bootstrap approach was used to estimate the Kappa CI [46]. For the rest of the metrics, the Clopper–Pearson method was applied [47].

3. Results and Discussion

#### 3.1. Recurrent Neural Networks (RNN)

Table 3 summarizes the results obtained using the LSTM model. The accuracy varied with the window size, with the best result (Acc = 89.53%) obtained for a window size of 25 ms.

Table 3. Performance achieved by the LSTM classifier. Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) with 95% confidence intervals (95% CI). The best accuracy is in bold.

Window	Acc	Pr	Se	Sp	FPR	Kappa
25 ms	**89.53** **(88.57–90.43)**	93.95 (92.89–94.9)	87.22 (85.83–88.52)	92.58 (91.29–93.74)	7.42 (6.26–8.71)	0.79 (0.77–0.80)
50 ms	89.23 (88.26–90.14)	93.60 (92.51–94.57)	87.00 (85.60–88.31)	92.16 (90.84–93.35)	7.84 (6.65–9.16)	0.78 (0.76–0.80)
70 ms	87.15 (86.11–88.14)	90.95 (89.70–92.10)	85.63 (84.16–87.00)	89.10 (87.61–90.47)	10.90 (9.53–12.39)	0.74 (0.72–0.76)
100 ms	87.17 (86.14–88.16)	90.16 (88.86–91.35)	86.20 (84.75–87.57)	88.37 (86.85–89.77)	11.63 (10.23–13.15)	0.74 (0.72–0.76)
200 ms	88.97 (87.99–89.89)	92.28 (91.10–93.34)	87.53 (86.13–88.83)	90.78 (89.39–92.05)	9.22 (7.95–10.61)	0.78 (0.76–0.80)
400 ms	88.20 (87.2–89.15)	93.07 (91.95–94.08)	85.80 (84.36–87.16)	91.42 (90.04–92.66)	8.58 (7.34–9.96)	0.76 (0.74–0.78)

3.2. Convolutional Neural Networks (CNN)

Figure 7 shows some examples of spectrograms of cough and non-cough signals. Table 4 summarizes the results obtained using the CNN model, when using the mel, ERB, and Bark spectrograms. Different window sizes were evaluated, and the best results were achieved for a window size of 25 ms. The accuracy and precision in the best case were 89.13% and 91.35% (ERB spectrogram).
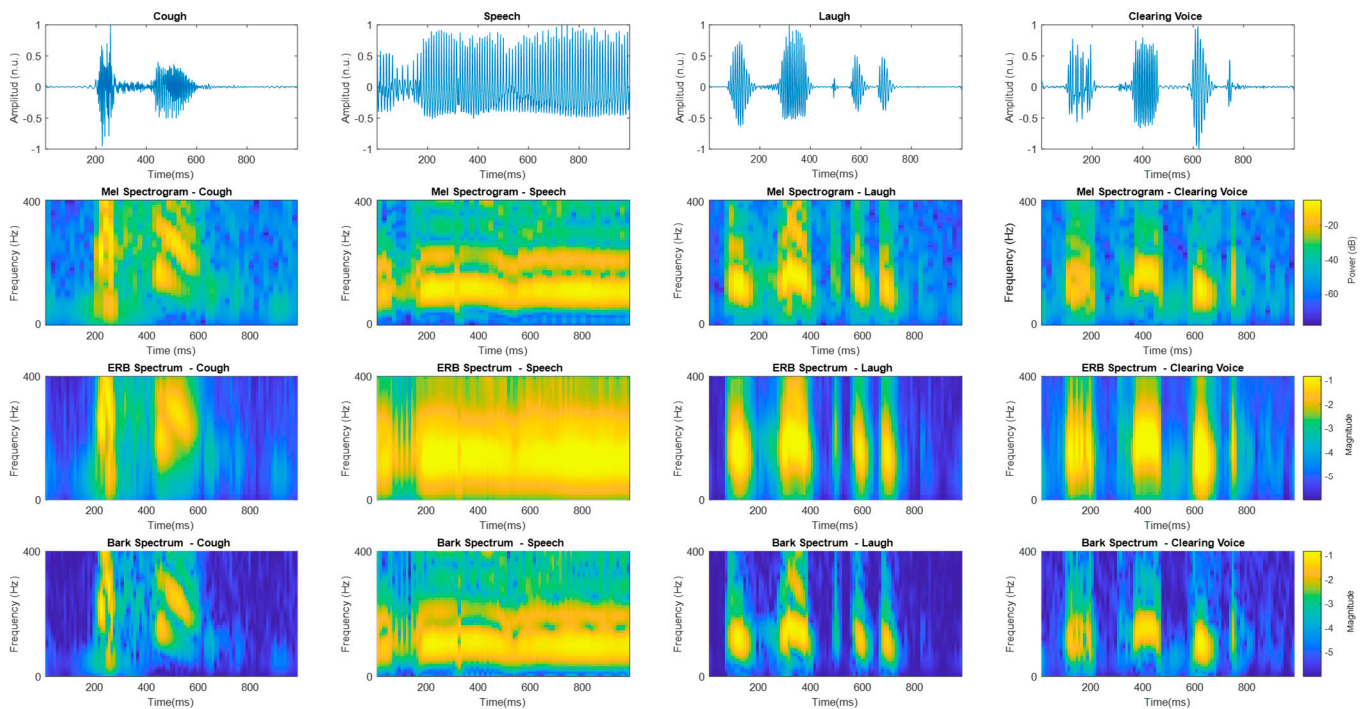
Figure 7. Examples of ERB, mel, and Bark spectrograms for cough and non-cough signals.

Table 4. Performance achieved by the CNN classifier using the proposed 2D feature extraction techniques (mel, ERB, and Bark spectrograms). Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI) are provided. The best accuracy is in bold.

Spectrogram	Acc	Pr	Se	Sp	FPR	Kappa
ERB	**89.13** **(88.16–90.05)**	91.35 (90.12–92.48)	88.46 (87.1–89.73)	89.94 (88.52–91.24)	10.06 (8.76–11.48)	0.78 (0.76–0.80)
Mel	88.62 (87.63–89.56)	90.25 (88.95–91.44)	88.45 (87.08–89.73)	88.82 (87.34–90.17)	11.18 (9.83–12.66)	0.77 (0.75–0.79)
Bark	88.41 (87.41–89.35)	91.13 (89.88–92.27)	87.46 (86.06–88.77)	89.57 (88.12–90.9)	10.43 (9.1–11.88)	0.77 (0.75–0.79)

3.3. Mixed RNN-CNN approach: VGGish

A VGGish pre-trained network was fine-tuned using transfer learning and fed with the log-mel spectrograms calculated using each acceleration signal segment. Table 5 summarizes the validation results. An accuracy of 90.35% was achieved.

Table 5. Performance achieved by the VGGish classifier. Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI) are provided.

Acc	Pr	Se	Sp	FPR	Kappa
90.35 (89.42–91.21)	95.28 (94.32–96.11)	87.55 (86.18–88.83)	94.13 (92.95–95.16)	5.87 (4.84–7.05)	80.52 (78.74–82.31)

3.4. Deep Convolutional Neural Networks (DCNNs)

Figure 8 shows the mel spectrogram, the kurtogram, the scalogram, and the signal envelope of some cough and non-cough signals. Tables 6–9 summarize the results ob-

tained using all the proposed DCNN models (ResNet-50, SqueezeNet, Inception ResNetV2, MobileNet-V2) and the proposed 2D features.

Table 6. Performance achieved by the DCNNs with mel spectrograms as input. Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI). The best accuracy is in bold.

Network	Acc	Pr	Se	Sp	FPR	Kappa
ResNet-50	89.41 (88.45–90.32)	91.35 (90.12–92.48)	88.92 (87.57–90.16)	90.00 (88.59–91.29)	10.00 (8.71–11.41)	0.78 (0.77–0.81)
SqueezeNet	**90.02 (89.08–90.90)**	92.85 (91.71–93.88)	88.78 (87.44–90.02)	91.55 (90.22–92.76)	8.45 (7.24–9.78)	0.80 (0.78–0.82)
Inception ResNetV2	89.58 (88.62–90.47)	92.28 (91.1–93.34)	88.49 (87.13–89.75)	90.91 (89.54–92.16)	9.09 (7.84–10.46)	0.79 (0.77–0.81)
MobileNet-V2	88.57 (87.58–89.51)	89.06 (87.70–90.31)	89.29 (87.94–90.54)	87.77 (86.27–89.17)	12.23 (10.83–13.73)	0.77 (0.75–0.79)

Table 7. Performance achieved by the deep DCNNs with the kurtogram as input. Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI). The best accuracy is in bold.

| Network	Acc	Pr	Se	Sp	FPR	Kappa
ResNet-50	79.10 (77.86–80.31)	86.94 (85.48–88.30)	76.65 (74.97–78.28)	82.77 (80.90–84.53)	17.23 (15.47–19.10)	0.58 (0.55–0.60)
SqueezeNet	79.29 (78.05–80.5)	90.07 (88.76–91.27)	75.48 (73.81–77.09)	85.8 (83.98–87.48)	14.2 (12.52–16.02)	0.58 (0.55–0.60)
Inception ResNetV2	**79.99 (78.76–81.18)**	90.95 (89.7–92.1)	75.94 (74.29–77.54)	86.98 (85.21–88.6)	13.02 (11.4–14.79)	0.59 (0.57–0.62)
MobileNet-V2	79.55 (78.31–80.75)	90.16 (88.86–91.35)	75.75 (74.09–77.36)	85.98 (84.18–87.65)	14.02 (12.35–15.82)	0.58 (0.56–0.61)

Table 8. Performance achieved by the DCNNs with the wavelet scalogram as input. Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI). The best accuracy is in bold.

Network	Acc	Pr	Se	Sp	FPR	Kappa
ResNet-50	90.35 (89.42–91.21)	93.47 (92.37–94.45)	88.84 (87.51–90.08)	92.23 (90.93–93.39)	7.77 (6.61–9.07)	0.81 (0.79–0.82)
SqueezeNet	**92.21 (91.37–93)**	95.59 (94.66–96.4)	90.25 (88.99–91.41)	94.7 (93.6–95.67)	5.3 (4.33–6.4)	0.84 (0.83–0.86)
Inception ResNetV2	90.21 (89.28–91.08)	91.35 (90.12–92.48)	90.24 (88.95–91.42)	90.17 (88.78–91.44)	9.83 (8.56–11.22)	0.80 (0.79–0.82)
MobileNet-V2	90.72 (89.81–91.57)	91.92 (90.73–93.01)	90.64 (89.38–91.8)	90.8 (89.45–92.04)	9.2 (7.96–10.55)	0.81 (0.8–0.83)

In the case of the mel spectrogram, the best result was achieved with the SqueezeNet model (accuracy of 90.02%). In the case of the kurtogram, the best result was obtained using the Inception ResnetV2 architecture, which led to an accuracy of 79.99%. The best performance using the scalogram was found in the SqueezeNet model, which reached an accuracy of 92.21%. Finally, the signal envelope provided the best result with the SqueezeNet model, with an accuracy of 89.67%

Table 9. Performance achieved by the DCNNs with the signal envelope as input. Acc: accuracy; Pr: precision; Se: sensitivity; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI). The best accuracy is in bold.

Network	Acc	Pr	Se	Sp	FPR	Kappa
ResNet-50	86.82 (85.77–87.82)	89.10 (87.74–90.35)	86.39 (84.94–87.76)	87.34 (85.78–88.78)	12.66 (11.22–14.22)	0.74 (0.71–0.76)
SqueezeNet	**89.67** **(88.72–90.56)**	93.65 (92.56–94.61)	87.65 (86.27–88.93)	92.29 (90.98–93.46)	7.71 (6.54–9.02)	0.79 (0.77–0.81)
Inception ResNetV2	88.25 (87.24–89.20)	89.19 (87.84–90.44)	88.64 (87.27–89.91)	87.80 (86.29–89.20)	12.20 (10.80–13.71)	0.76 (0.74–0.78)
MobileNet-V2	87.27 (86.23–88.25)	88.48 (87.1–89.77)	87.55 (86.13–88.88)	86.94 (85.38–88.38)	13.06 (11.62–14.62)	0.74 (0.72–0.76)
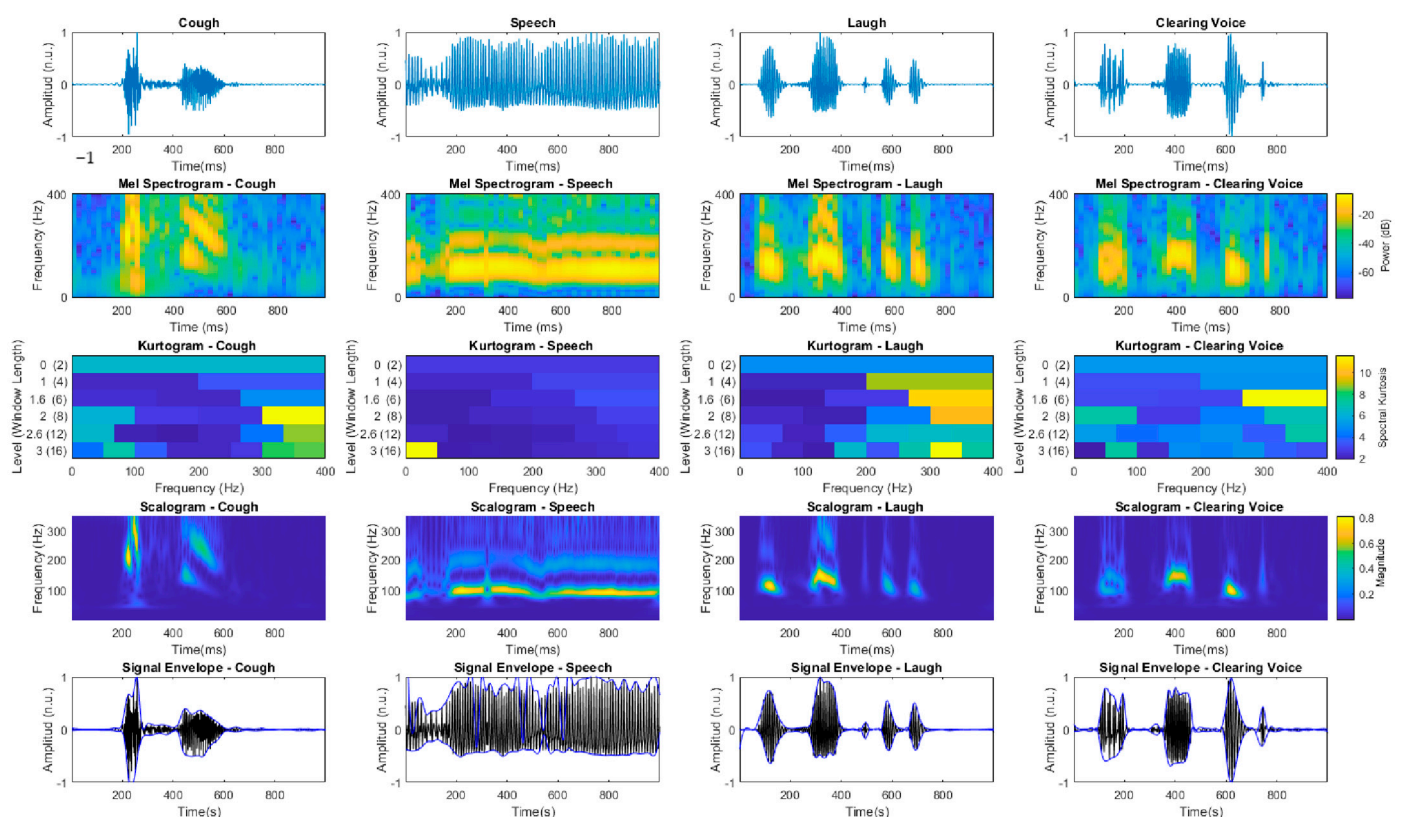
Figure 8. Examples of mel, kurtogram, wavelet scalogram, and signal envelope for cough and non-cough signals.

Table 10 summarizes the best results obtained among all the DL models evaluated. As shown, mel and ERB spectrograms, as well as wavelet scalograms combined with CNN, DCNN, and mixed CNN + RNN models, operated with high performance in the classification of cough events, discriminating them with a high degree of efficiency from other sounds such as speech, laugh, or voice-clearing noises.

In particular, the SqueezeNet model with wavelet scalograms with an accuracy of 92.21% and the VGGish architecture, which used mel spectrograms, with an accuracy of 90.35%, outperformed the other models. The performance of the CNN and RNN models was also noteworthy, with accuracies greater than 89%.

Table 10. Summary of the best results. Acc: accuracy; Pr: precision; Sp: specificity; FPR: false positive rate. All metrics are expressed as percentages (%) and 95% confidence intervals (95% CI). The best accuracy is in bold.

Model	Features	Acc	Pr	Sp	FPR
RNN	1D	89.53 (88.57–90.43)	93.95 (92.89–94.9)	92.58 (91.29–93.74)	7.42 (6.26–8.71)
CNN-RNN-VGGish	2D–Mel spectrogram	90.35 (89.42–91.21)	95.28 (94.32–96.11)	94.13 (92.95–95.16)	5.87 (4.84–7.05)
CNN	2D–ERB spectrogram	89.13 (88.16–90.05)	91.35 (90.12–92.48)	89.94 (88.52–91.24)	10.06 (8.76–11.48)
DCNN-SqueezeNet	**2D–Wavelet scalogram**	**92.21 (91.37–93)**	95.59 (94.66–96.4)	94.70 (93.6–95.67)	5.30 (4.33–6.40)

4. Discussion

In this work, we presented an approach for the automatic classification of cough events using a single sensor single signal system. Deep learning techniques were applied to the signal recorded by a digital accelerometer sensor located in the suprasternal notch. One- and two-dimensional time–frequency features were thoroughly explored.

Although cough assessment has been carried out conventionally using microphones [48], the discrimination of cough and non-cough activities remains a challenge [49]. For this purpose, the use of accelerometers has been proposed as a robust alternative. The use of accelerometers for the quantification of biomedical signals such as heart rate and variability, respiratory rate, or snoring is widespread in the scientific literature and presents certain benefits over conventional microphones [50]. The sampling rate is significantly lower, which substantially reduces the computational burden in most DL models. This enables the use of complex deep learning models that can be deployed on portable devices. In addition, sensitivity to external background noise is significantly reduced, given the coupling of the sensor to the patient's body. Finally, privacy is enhanced since the continuous monitoring does not involve recording audio signals.

However, unlike the processing of sounds recorded by microphones to detect coughing, the use of current deep learning techniques applied to signals recorded by acceleration sensors to detect cough has been little explored. In [51], an accelerometer system was used to evaluate a self-contained, ambulatory cough monitor using conventional signal processing techniques. More recently, the use of deep learning techniques together with accelerometry signals from a smartphone placed at the patient's bedside has been assessed [52]. Bed-mounted sensors represent a good option for monitoring nocturnal cough but are not valid for assessing cough in ambulatory conditions. In [49], a dual accelerometry was placed on the patient's neck, and 1D temporal, time–frequency, frequency, and information-theoretic features were used with artificial neural networks and support vector machines, achieving an accuracy of 90.2% in detecting voluntary cough.

In our study, a single-axis signal from an accelerometer placed on the patient's suprasternal notch was used to screen cough. DL models of four architectures were evaluated, and the performance of 1D and 2D time–frequency representations used as input to these models was thoroughly assessed. Models that offer a variety of approaches to cough signal classification were explored, from recurrent networks capturing temporal sequences, to deep and lightweight convolutional networks extracting relevant features from signal spectrograms.

Overall, all the DL architectures and 1D/2D representations evaluated showed high performance. The DCNN models using the kurtogram as input showed the lowest efficiency. The kurtogram is a useful tool for detecting signal transients, but several aspects could limit its effectiveness in the specific application for cough detection using acceleration signals. The kurtogram offers a particular resolution that may not be best suited for all types of transients present in a cough. It appears that this resolution does not adequately capture the distinctive features of the cough.

However, the 1D spectral and temporal features showed discriminative ability with the LSTM model (RNN). The same was observed for the ERB, mel, and Bark spectrograms with CNN networks, which allow a better representation of transients and frequency components relevant to the coughing process. These 1D and 2D representations can provide a detailed plot in both the time and frequency domain, which is crucial for capturing the transient characteristics of a cough.

The signal envelope captures the slow variation of the signal amplitude, eliminating the high-frequency components. This is particularly useful for identifying high-energy events such as coughing but does not provide the same detailed information in the frequency domain as spectrograms, which may hinder its accuracy. However, both the scalograms and the coefficients derived from mel spectrograms were found to be particularly effective. Mel spectrograms are widely used in pattern recognition and deep learning due to their effectiveness in extracting relevant features for analyzing acoustic signals. These scales apply nonlinear transformations that may be more effective in capturing the discriminative characteristics of complex signals such as acceleration signals linked to coughing. Scalograms use wavelet transforms. Unlike other transforms, wavelets allow better resolution in both domains, efficiently capturing both transients and stationary components, which may underlie the success of models that used these 2D features.

The selection of the optimal model depends on factors such as the required accuracy, available computational resources, and deployment context. The SqueezeNet model with wavelet scalograms and the VGGish model with mel spectrograms performed consistently well, with high accuracy (92.21% and 90.35%, respectively) and specificity (94.70% and 94.13% respectively). Both models produced a low false positive ratio (lower than 6%), indicating high robustness to overcome the main barrier in cough detection (false positives due to background noise or other similar sounds such as voice, laughing, or throat clearing).

The performance of the SqueezeNet model was outstanding. It is a compact architecture designed to be efficient in terms of model size and computational requirements while maintaining high performance in classification tasks. It can make inferences faster, which is crucial for real-time applications in devices with limited computational performance (e.g., smartphones, embedded systems, and wearables).

Finally, this study presents limitations. Firstly, a larger sample size is required to validate the results. Secondly, voluntary coughs present different patterns from spontaneous coughs [53]. However, the ability to detect voluntary coughs under acoustically challenging ambient conditions is a first step toward a clinically applicable ambulatory cough monitoring system [54]. Consequently, since the cough episodes used in this study were collected from healthy subjects and were simulated, future work must focus on validating the system in spontaneous coughing of subjects with respiratory diseases under natural ambulatory conditions. In addition, the next steps include the discrimination of wet and dry cough and the study of gender differences in cough patterns.

5. Conclusions

In this study, we evaluated a cough detection system using a digital accelerometer to capture single-axis acceleration signals. The system's ability to distinguish voluntary coughs from background noises was extensively tested using DL techniques and time–frequency representations. The most effective approaches involved scalograms and mel spectrograms and were achieved with the SqueezeNet and the VGGish models, respectively. The results obtained show a promising potential for cough monitoring.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Irwin, R.S.; French, C.T.; Lewis, S.Z.; Diekemper, R.L.; Gold, P.M.; Adams, T.M.; Altman, K.W.; Barker, A.F.; Birring, S.S.; Bolser, D.C.; et al. Overview of the Management of Cough: CHEST Guideline and Expert Panel Report. *Chest* **2014**, *146*, 885–889. [CrossRef]
2. Song, W.J.; Chang, Y.S.; Faruqi, S.; Kim, J.Y.; Kang, M.G.; Kim, S.; Jo, E.J.; Kim, M.H.; Plevkova, J.; Park, H.W.; et al. The Global Epidemiology of Chronic Cough in Adults: A Systematic Review and Meta-Analysis. *Eur. Respir. J.* **2015**, *45*, 1479–1481. [CrossRef]
3. Morice, A.H.; Fontana, G.A.; Belvisi, M.G.; Birring, S.S.; Chung, K.F.; Dicpinigaitis, P.V.; Kastelik, J.A.; McGarvey, L.P.; Smith, J.A.; Tatar, M.; et al. ERS Guidelines on the Assessment of Cough. *Eur. Respir. J.* **2007**, *29*, 1256–1276. [CrossRef]
4. Fietze, I.; Dingli, K.; Diefenbach, K.; Douglas, N.J.; Glos, M.; Tallafuss, M.; Terhalle, W.; Witt, C. Night-to-Night Variation of the Oxygen Desaturation Index in Sleep Apnoea Syndrome. *Eur. Respir. J.* **2004**, *24*, 987–993. [CrossRef]
5. Gallo Marin, B.; Aghagoli, G.; Lavine, K.; Yang, L.; Siff, E.J.; Chiang, S.S.; Salazar-Mather, T.P.; Dumenco, L.; Savaria, M.C.; Aung, S.N.; et al. Predictors of COVID-19 Severity: A Literature Review. *Rev. Med. Virol.* **2020**, *31*, 1–10. [CrossRef]
6. Ekberg-Aronsson, M.; Pehrsson, K.; Nilsson, J.Å.; Nilsson, P.M.; Löfdahl, C.G. Mortality in GOLD Stages of COPD and Its Dependence on Symptoms of Chronic Bronchitis. *Respir. Res.* **2005**, *6*, 98. [CrossRef]
7. Brignall, K.; Jayaraman, B.; Birring, S.S. Quality of Life and Psychosocial Aspects of Cough. *Lung* **2008**, *186*, 55–58. [CrossRef]
8. Birring, S.S. New Concepts in the Management of Chronic Cough. *Pulm. Pharmacol. Ther.* **2011**, *24*, 334–338. [CrossRef]
9. Hall, J.I.; Lozano, M.; Estrada-Petrocelli, L.; Birring, S.; Turner, R. The Present and Future of Cough Counting Tools. *J. Thorac. Dis.* **2020**, *12*, 5207–5223. [CrossRef]
10. Song, W.-J.; Dupont, L.; Birring, S.S.; Chung, K.F.; browska, M.D.; Dicpinigaitis, P.; Ribas, C.D.; Fontana, G.; Gibson, P.G.; Guilleminault, L.; et al. Early View Consensus Goals and Standards for Specialist Cough Clinics: The NEUROCOUGH International Delphi Study. *ERJ Open Res.* **2023**, *9*, 00618–2023. [CrossRef]
11. Hirons, B.; Rhatigan, K.; Kesavan, H.; Turner, R.D.; Birring, S.S.; Cho, P.S.P. Cough in Chronic Lung Disease: A State of the Art Review. *J. Thorac. Dis.* **2023**, *15*, 5823–5843. [CrossRef]
12. Marsden, P.A.; Smith, J.A.; Kelsall, A.A.; Owen, E.; Naylor, J.R.; Webster, D.; Sumner, H.; Alam, U.; McGuinness, K.; Woodcock, A.A. A Comparison of Objective and Subjective Measures of Cough in Asthma. *J. Allergy Clin. Immunol.* **2008**, *122*, 903–907. [CrossRef]
13. Mei, H.; Gu, W.; Ran, L.; Wen, S.; Yu, L.; Xu, X. Evaluation Methods and Influencing Factors of Cough Sensitivity. *Ther. Adv. Respir. Dis.* **2022**, *16*, 175346662110701. [CrossRef]
14. Steele, A.M.; Meuret, A.E.; Millard, M.W.; Ritz, T. Discrepancies between Lung Function and Asthma Control: Asthma Perception and Association with Demographics and Anxiety. *Allergy Asthma Proc.* **2012**, *33*, 500–507. [CrossRef]
15. Smith, J.A.; Woodcock, A. Cough Recording Technology. *Curr. Respir. Med. Rev.* **2011**, *7*, 34–39. [CrossRef]
16. McGuinness, K.; Holt, K.; Dockry, R.; Smith, J. P159 Validation of the VitaloJAK^{TM} 24 Hour Ambulatory Cough Monitor: Abstract P159 Table 1. *Thorax* **2012**, *67*, A131. [CrossRef]
17. Barry, S.J.; Dane, A.D.; Morice, A.H.; Walmsley, A.D. The Automatic Recognition and Counting of Cough. *Cough* **2006**, *2*, 8. [CrossRef]
18. Kvapilova, L.; Boza, V.; Dubec, P.; Majernik, M.; Bogar, J.; Jamison, J.; Goldsack, J.C.; Kimmel, D.J.; Karlin, D.R. Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough. *Digit. Biomark.* **2019**, *3*, 166–175. [CrossRef]
19. Yang, X.; Zhang, M.; Gong, S.; Sun, M.; Xie, M.; Niu, P.; Pang, W. Chest-Laminated and Sweat-Permeable E-Skin for Speech and Motion Artifact-Insensitive Cough Detection. *Adv. Mater. Technol.* **2023**, *8*, 2201043. [CrossRef]
20. Pahar, M.; Miranda, I.; Diacon, A.; Niesler, T. Accelerometer-Based Bed Occupancy Detection for Automatic, Non-Invasive Long-Term Cough Monitoring. *IEEE Access* **2023**, *11*, 30739–30752. [CrossRef]
21. Pahar, M.; Miranda, I.; Diacon, A.; Niesler, T. Deep Neural Network Based Cough Detection Using Bed-Mounted Accelerometer Measurements. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; IEEE: Piscataway, NJ, USA, 2021; Volume 2021, pp. 8002–8006.
22. Otoshi, T.; Nagano, T.; Izumi, S.; Hazama, D.; Katsurada, N.; Yamamoto, M.; Tachihara, M.; Kobayashi, K.; Nishimura, Y. A Novel Automatic Cough Frequency Monitoring System Combining a Triaxial Accelerometer and a Stretchable Strain Sensor. *Sci. Rep.* **2021**, *11*, 9973. [CrossRef]
23. Drugman, T.; Urbain, J.; Bauwens, N.; Chessini, R.; Valderrama, C.; Lebecque, P.; Dutoit, T. Objective Study of Sensor Relevance for Automatic Cough Detection. *IEEE J. Biomed. Health Inf.* **2013**, *17*, 699–707. [CrossRef]
24. Ijaz, A.; Nabeel, M.; Masood, U.; Mahmood, T.; Hashmi, M.S.; Posokhova, I.; Rizwan, A.; Imran, A. Towards Using Cough for Respiratory Disease Diagnosis by Leveraging Artificial Intelligence: A Survey. *Inf. Med. Unlocked* **2022**, *29*, 100832. [CrossRef]

25. Rendon, D.B.; Ojeda, J.L.R.; Foix, L.F.C.; Morillo, D.S.; Fernandez, M.A. Mapping the Human Body for Vibrations Using an Accelerometer. In Proceedings of the Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, Lyon, France, 22–26 August 2007; pp. 1671–1674.

26. Smith, J.; Woodcock, A. New Developments in the Objective Assessment of Cough. *Lung* **2008**, *186* (Suppl. S1), 48–54. [CrossRef]

27. Zhou, Q.; Shan, J.; Ding, W.; Wang, C.; Yuan, S.; Sun, F.; Li, H.; Fang, B. Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network. *Front. Robot. AI* **2021**, *8*, 580080. [CrossRef]

28. Egger, J.; Zhang, C.-X.; Jo, A.-H.; Kwak, K.-C. Speech Emotion Recognition Based on Two-Stream Deep Learning Model Using Korean Audio Information. *Appl. Sci.* **2023**, *13*, 2167. [CrossRef]

29. Chabot, P.; Bouserhal, R.E.; Cardinal, P.; Voix, J. Detection and Classification of Human-Produced Nonverbal Audio Events. *Appl. Acoust.* **2021**, *171*, 107643. [CrossRef]

30. Kulevome, D.K.B.; Wang, H.; Cobbinah, B.M.; Mawuli, E.S.; Kumar, R. Effective Time-Series Data Augmentation with Analytic Wavelets for Bearing Fault Diagnosis. *Expert Syst. Appl.* **2024**, *249*, 123536. [CrossRef]

31. Askari Nasab, K.; Mirzaei, J.; Zali, A.; Gholizadeh, S.; Akhlaghdoust, M. Coronavirus Diagnosis Using Cough Sounds: Artificial Intelligence Approaches. *Front. Artif. Intell.* **2023**, *6*, 1100112. [CrossRef]

32. Alías, F.; Socoró, J.C.; Sevillano, X. A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Appl. Sci.* **2016**, *6*, 143. [CrossRef]

33. Kim, H.; Moreau, N.; Sikora, T. *MPEG-7 Audio and Beyond*; Wiley: Hoboken, NJ, USA, 2005; ISBN 9780470093344.

34. Srivastava, A.; Jain, S.; Miranda, R.; Patil, S.; Pandya, S.; Kotecha, K. Deep Learning Based Respiratory Sound Analysis for Detection of Chronic Obstructive Pulmonary Disease. *PeerJ Comput. Sci.* **2021**, *7*, e369. [CrossRef]

35. Palaniappan, R.; Sundaraj, K.; Ahamed, N.U. Machine Learning in Lung Sound Analysis: A Systematic Review. *Biocybern. Biomed. Eng.* **2013**, *33*, 129–135. [CrossRef]

36. Alqudaihi, K.S.; Aslam, N.; Khan, I.U.; Almuhaideb, A.M.; Alsunaidi, S.J.; Ibrahim, N.M.A.R.; Alhaidari, F.A.; Shaikh, F.S.; Alsenbel, Y.M.; Alalharith, D.M.; et al. Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities. *IEEE Access* **2021**, *9*, 102327–102344. [CrossRef]

37. Kim, Y.; Hyon, Y.K.; Jung, S.S.; Lee, S.; Yoo, G.; Chung, C.; Ha, T. Respiratory Sound Classification for Crackles, Wheezes, and Rhonchi in the Clinical Field Using Deep Learning. *Sci. Rep.* **2021**, *11*, 17186. [CrossRef]

38. Syed, S.A.; Rashid, M.; Hussain, S.; Zahid, H. Comparative Analysis of CNN and RNN for Voice Pathology Detection. *BioMed Res. Int.* **2021**, *2021*, 6635964. [CrossRef]

39. Hamdi, S.; Oussalah, M.; Moussaoui, A.; Saidi, M. Attention-Based Hybrid CNN-LSTM and Spectral Data Augmentation for COVID-19 Diagnosis from Cough Sound. *J. Intell. Inf. Syst.* **2022**, *59*, 367–389. [CrossRef]

40. Pahar, M.; Klopper, M.; Reeve, B.; Warren, R.; Theron, G.; Diacon, A.; Niesler, T. Automatic Tuberculosis and COVID-19 Cough Classification Using Deep Learning. In Proceedings of the 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 20–22 July 2022. [CrossRef]

41. Sharan, R.V.; Rahimi-Ardabili, H. Detecting Acute Respiratory Diseases in the Pediatric Population Using Cough Sound Features and Machine Learning: A Systematic Review. *Int. J. Med. Inf.* **2023**, *176*, 105093. [CrossRef]

42. Nguyen, M.T.; Lin, W.W.; Huang, J.H. Heart Sound Classification Using Deep Learning Techniques Based on Log-Mel Spectrogram. *Circuits Syst. Signal Process.* **2023**, *42*, 344–360. [CrossRef]

43. Salamea-Palacios, C.R.; Sanchez-Almeida, T.; Calderon-Hinojosa, X.; Guana-Moya, J.; Castaneda-Romero, P.; Reina-Travez, J. On the Use of VGGish as Feature Extractor for COVID-19 Cough Classification. In *Proceedings of the 2023 8th International Conference on Machine Learning Technologies*; Association for Computing Machinery: New York, NY, USA, 2023; pp. 89–94. [CrossRef]

44. Aly, M.; Alotaibi, N.S. A Novel Deep Learning Model to Detect COVID-19 Based on Wavelet Features Extracted from Mel-Scale Spectrogram of Patients' Cough and Breathing Sounds. *Inf. Med. Unlocked* **2022**, *32*, 101049. [CrossRef]

45. Dentamaro, V.; Giglio, P.; Impedovo, D.; Moretti, L.; Pirlo, G. AUCO ResNet: An End-to-End Network for COVID-19 Pre-Screening from Cough and Breath. *Pattern Recognit.* **2022**, *127*, 108656. [CrossRef]

46. Fung, K.P.; Lee, J. Bootstrap Estimate of the Variance and Confidence Interval of Kappa. *Br. J. Ind. Med.* **1991**, *48*, 503. [CrossRef] [PubMed]

47. Clopper, C.J.; Pearson, E.S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **1934**, *26*, 404–413. [CrossRef]

48. Birring, S.S.; Fleming, T.; Matos, S.; Raj, A.A.; Evans, D.H.; Pavord, I.D. The Leicester Cough Monitor: Preliminary Validation of an Automated Cough Detection System in Chronic Cough. *Eur. Respir. J.* **2008**, *31*, 1013–1018. [CrossRef] [PubMed]

49. Mohammadi, H.; Samadani, A.A.; Steele, C.; Chau, T. Automatic Discrimination between Cough and Non-Cough Accelerometry Signal Artefacts. *Biomed. Signal Process. Control* **2019**, *52*, 394–402. [CrossRef]

50. Morillo, D.S.; Rojas Ojeda, J.L.; Crespo Foix, L.F.; Jiménez, A.L. An Accelerometer-Based Device for Sleep Apnea Screening. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 491–499. [CrossRef] [PubMed]

51. Paul, I.M.; Wai, K.; Jewell, S.J.; Shaffer, M.L.; Varadan, V.V. Evaluation of a New Self-Contained, Ambulatory, Objective Cough Monitor. *Cough* **2006**, *2*, 7. [CrossRef] [PubMed]

52. Pahar, M.; Miranda, I.; Diacon, A.; Niesler, T. Automatic Non-Invasive Cough Detection Based on Accelerometer and Audio Signals. *J. Signal Process. Syst.* **2022**, *94*, 821. [CrossRef]

53. Smith, J.A.; Ashurst, L.; Jack, S.; Woodcock, A.A.; Earis, J.E. The Description of Cough Sounds by Healthcare Professionals. *Cough* **2006**, *2*, 1. [CrossRef]

54. Vizel, E.; Yigla, M.; Goryachev, Y.; Dekel, E.; Felis, V.; Levi, H.; Kroin, I.; Godfrey, S.; Gavriely, N. Validation of an Ambulatory Cough Detection and Counting Application Using Voluntary Cough under Different Conditions. *Cough* **2010**, *6*, 3. [CrossRef]