Article

Classification Predictive Model for Air Leak Detection in Endoworm Enteroscopy System

Roberto Zazo-Manzaneque [1,*], Vicente Pons-Beltrán [2,3], Ana Vidaurre [1,4], Alberto Santonja [5] and Carlos Sánchez-Díaz [6]

1 Centre for Biomaterials and Tissue Engineering (CBIT), Universitat Politècnica de València, 46022 Valencia, Spain; vidaurre@fis.upv.es
2 Digestive Endoscopy Unit, Digestive Diseases Department, La Fe Polytechnic Univesity Hospital, 46026 Valencia, Spain; pons_vicbel@gva.es
3 Gastrointestinal Endoscopy Research Group, IIS Hospital La Fe, 46026 Valencia, Spain
4 Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), 28029 Madrid, Spain
5 School of Design Engineering (ETSID), Universitat Politècnica de València, 46022 Valencia, Spain; asantonj@upvnet.upv.es
6 Department of Electronic Engineering, Universitat Politècnica de València, 46022 Valencia, Spain; csanched@eln.upv.es
* Correspondence: rozaman@doctor.upv.es

Abstract: Current enteroscopy techniques present complications that are intended to be improved with the development of a new semi-automatic device called Endoworm. It consists of two different types of inflatable cavities. For its correct operation, it is essential to detect in real time if the inflatable cavities are malfunctioning (presence of air leakage). Two classification predictive models were obtained, one for each cavity typology, which must discern between the "*Right*" or "*Leak*" states. The cavity pressure signals were digitally processed, from which a set of features were extracted and selected. The predictive models were obtained from the features, and a prior classification of the signals between the two possible states was used as input to different supervised machine learning algorithms. The accuracy obtained from the classification predictive model for cavities of the *balloon-type* was 99.62%, while that of the *bellows-type* was 100%, representing an encouraging result. Once the models are validated with data generated in animal model tests and subsequently in exploratory clinical tests, their incorporation in the software device will ensure patient safety during small bowel exploration.

Keywords: classification predictive models; digital signal processing; enteroscopy; feature extraction; inflatable cavities; medical device; real-time detection system; soft robot

1. Introduction

Currently, there are different techniques for the exploration, diagnosis, and therapy of small bowel pathologies, the main one being enteroscopy. There are three different types of commercially available enteroscopes: single-balloon (SBE) [1], double-balloon (DBE) [2,3], and spiral (SE) [4]. Although these systems allow exploration of the small intestine, they are not without limitations [5–9].

In this context, a research group from the *Universitat Politècnica de València* and the *Fundación de Investigación del Hospital La Fe de Valencia* is working on the development of a new enteroscopy system, called Endoworm, which aims to improve the existing systems.

Endoworm is a semiautomatic soft robot device [10–13], which is mounted on a conventional endoscope and allows exploration of the small intestine. It consists of a pneumatic translation system of inflatable cavities governed by a microcontroller-based electronic device. The objective is to retract the intestine over the endoscope, thus assisting in advancing the endoscope [14,15].

The commercial systems used to explore the small bowel (especially DBE and SBE) have been tested for years, and some problems have been detected [16]; among them, no air leakage has been reported. The main reason for that is the manual operation of these two systems that allows the specialist to detect any malfunctioning of every inflatable cavity. Due to the automated operation of the Endoworm system and its complexity compared to the mentioned systems, the detection of malfunctioning in one of the three cavities is very difficult for the specialist, even more so to determine which one has failed. For this reason, developing an autonomous system for detecting air leakage from its inflatable cavities is necessary. An air leak in any of the cavities would represent a potential risk to the patient and even a significant loss in device efficiency to help the enteroscope advance through the small bowel.

Automated leak detection has been studied for different industrial applications. Leaks in rigid pipes have been investigated [17], some of them using machine learning methods [18]. In [19], a pneumatic system was analyzed to detect air leakage in the pipes and the pneumatic actuators while the system continues working. However, all the systems were rigid. Different methods to determine the leakage were presented in [20]. One of them used the pressure drop in the pipeline, again considering rigid pipes in industrial applications.

The complexity of detecting air leaks in Endoworm cavities lies in two aspects: their continuous inflation and deflation, which makes the static analysis of pressures ineffective, and the impossibility of seeing what is happening inside the patient, as well as the fact that it is not intended to introduce any electrical element that could cause damage. As described, leakage problems have been studied for industrial applications, but no previous studies have addressed this specific problem.

In medical applications, classification predictive models are commonly used to solve binary new cases classification, with high performance ratios [21]. These techniques can be used to classify the state of each cavity as "*Right*" or "*Leak*" while the device is working.

This work aimed to obtain two classification predictive models capable of detecting air leaks in the two different types of cavities (*balloon* and *bellows*) that make up the Endoworm translation system. The data used to generate the models were obtained by performing tests on the Endoworm enteroscopy system in in vitro models. The training and validation of the predictive models were carried out using *k-fold cross-validation* and test performance techniques. Finally, the best classification predictive model for each cavity type was selected, considering that they will have to be run in real time on the Endoworm control device.

2. Materials and Methods

Classification predictive models, which are intended to be obtained, require a series of input variables (features) that contain the necessary information on the pressure signals of the Endoworm cavities to classify the cases as "*Right*" or "*Leak*".

The starting point to obtain the features was the differential pressure signals (relative to atmospheric pressure) received from the sensors arranged in the air outlet ports of the Endoworm control device. There are four signals in total: system pressure, pressure in the two radial expansion cavities (*balloons*), and the pressure in the axial expansion cavity (*bellows*). A detailed description of the Endoworm control device, the three inflatable cavities, and the inflation-deflating sequence can be found in [15].

The system pressure typically ranges from 200 to 300 kPa, depending on the configuration entered by the user. The pressure signals, measured in real time from the *balloons* and *bellows*, vary from 0 kPa to system pressure.

Of these four pressure signals, only the three corresponding to inflatable cavities are of interest for air leak detection. Therefore, these will serve as sources of information from which to extract the features to be used as inputs for the predictive model of each type of cavity (*balloons* or *bellows*).

The cavities are directly connected to the "*Honeywell*" pressure sensor "*26PCFFA2G*", which can measure maximum differential pressure of 100 psi (689.476 kPa) with a sensitivity of 1 mV/psi (145.038 μV/kPa). This is important as the sensor model was changed from

that used in [14,15], increasing the full scale of the measurement to 320 kPa. The gain of the AD620 instrumentation operational amplifier was readjusted to 107.73 V/V, achieving the desired full scale and a resolution of 0.313 kPa/bit. Between the output of the AD620 and the ADC input of the *PIC18F4550* microcontroller, an anti-aliasing filter (first-order passive low pass filter) was placed. The cutoff frequency of the anti-aliasing filter is 10 Hz. This frequency was selected because all the representative spectral content of the signals is contained below this cutting frequency. Moreover, it allows a reliable representation in the time domain of the fast transients present in the signals. The sampling frequency selected for the signals is 200 Hz, resulting in a time resolution of 5 ms per sample.

An external module was developed to acquire the main analogue and digital signals, in real time, that define the behavior of the control device (*Sniffer*), avoiding overloading the device's control microcontroller.

The *Sniffer* was a board based on the *Atmega2560* microcontroller. It was connected via a matching board to the microcontroller pins of the control device. The digital I/O of the *Sniffer* and the control device were connected directly. In contrast, the analogue signals from the pressure sensors were connected to the *Sniffer* via four *rail-to-rail* operational amplifiers in buffer configuration (*MCP6044)*, whose function was impedance matching between the inputs to the ADCs of the *PIC18F4550* and the *Atmega2560*. Figure 1 shows a block diagram illustrating the connection of the *Sniffer* to the control device.
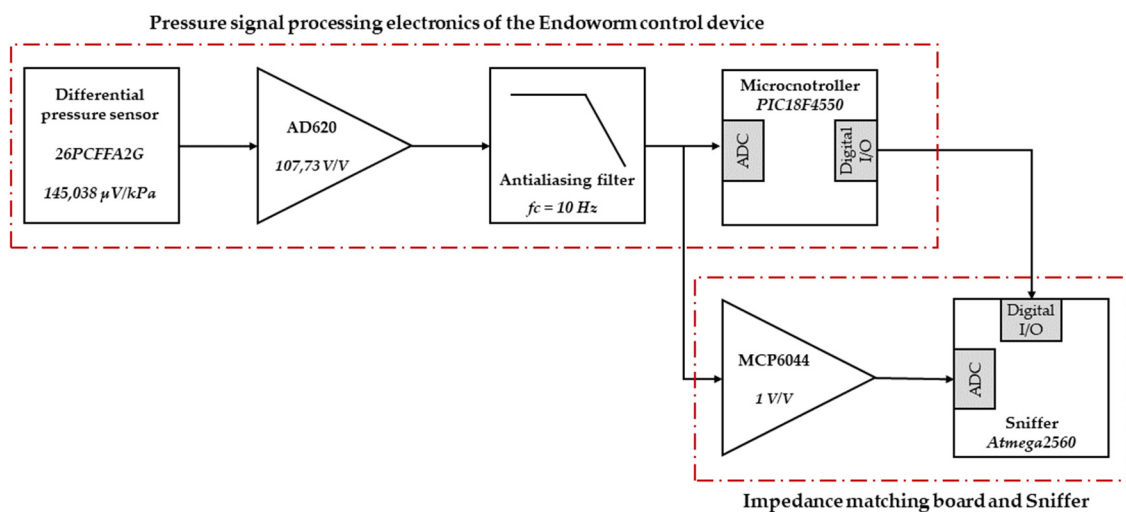

Figure 1. Diagram of the circuit for adapting the *26PCFFA2G* sensor output signal to the ADC input of the Endoworm microcontroller (*PIC18F4550*), as well as the capture of the pressure signals, through the impedance matching board, and the main digital control signals.

The *Sniffer* records and sends to an external PC the pressure signals, the states of the solenoid valves, the state of the pneumatic pumps, and the main parameters of the *Finite State Machine*, which determine the control behavior of the device. The theoretical basis on which the programming of the Endoworm control device is based can be found in [22].

The data used to generate the predictive models were obtained by performing tests on the Endoworm enteroscopy system, aiming to capture the device's normal functioning. A PU (polyurethane) artificial bowel model with an internal diameter of 40 mm was used for all tests. The intestine and the cavities used were characterized, and the results can be found in [14].

A total of 121 recordings were made, attempting to capture as much behavioral variability as possible in the operation of the Endoworm system. In some trials, the device was allowed to move freely (without being operated by any user); in others, it was driven by an expert endoscopist. The speed of the cavity sequence, its inflation–deflation times, and the system's maximum pressure were varied. Three possible values for the sequence speed ("*Low*", "*Medium*", and "*High*") could be established. The inflation and deflation

times determined by the opening times of the solenoid valves varied from 0.050 to 0.125 s
for *balloons* and from 0.8 to 1.4 s for *bellows*.

The "*Right*" or "*Leak*" operation was determined by the persons performing the
experiment, on the basis of the observation of air leaks, for each inflation–deflating cycle of
each cavity.

Figure 2 shows the pressure signals from the Endoworm control device corresponding
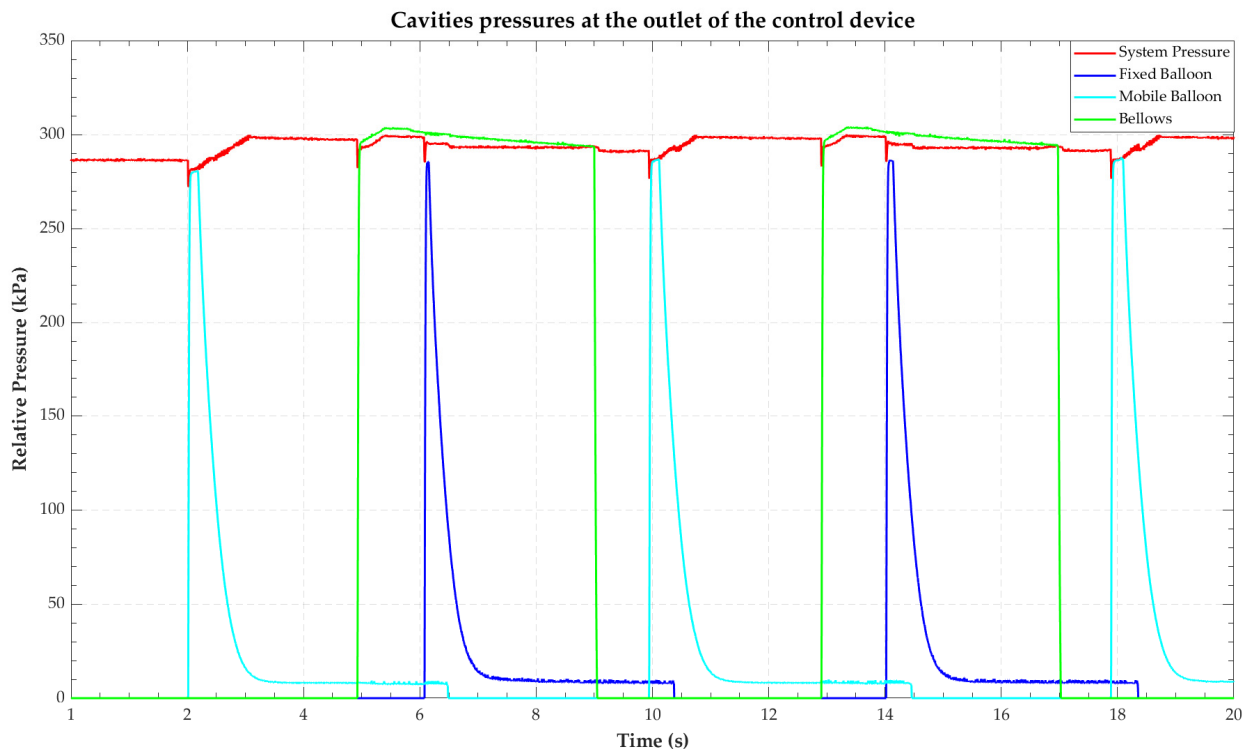to a segment of a recording in which all cavities were functioning correctly.

Figure 2. Temporal representation of the four pressures recorded by the *Sniffer* for any given test:
system pressure (red), fixed balloon (dark blue), mobile balloon (blue), and bellows (green).

The digital processing of the signals and the training of the classification predictive
models were carried out using the "*Matlab R2022a*" software under a research license
acquired by the *Universitat Poltècnica de València.*

A segmentation of the signals was carried out, consisting of two steps: signal swell
detection and windowing of the signal. Swell detection was set as the instant at which
the rising edge of the cavity pressure signal occurs. It was a robust parameter because
whenever the solenoid valve was ordered to open, an abrupt rise in pressure was recorded
in the sensors and calculated by detecting the positive peak of the derivative of the signal.
From the time at which inflation occurs, a signal windowing was established to cover a
complete cavity inflation–deflation cycle. This windowing was different depending on the
type of cavity.

Figure 3 shows an example of the segmentation of each type of cavity, revealing the
pressure signal, its derivative, and the instant in which the swelling of the cavity was
detected. In addition, the time window corresponding to the segmented signal is shown.

A total of 2350 *balloon-type* segmented signals and 924 *bellows-type* segmented signals
were obtained. These were assigned the label ("*Right*" or "*Leak*") designated by the qualified
personnel during the tests. Figure 4 shows typical signals representing the two possible
states of the two cavity types that make up the Endoworm.

Once the *balloon* and *bellows* signals were segmented and associated with their corre-
sponding classification, they were randomized. The data for each cavity type were divided
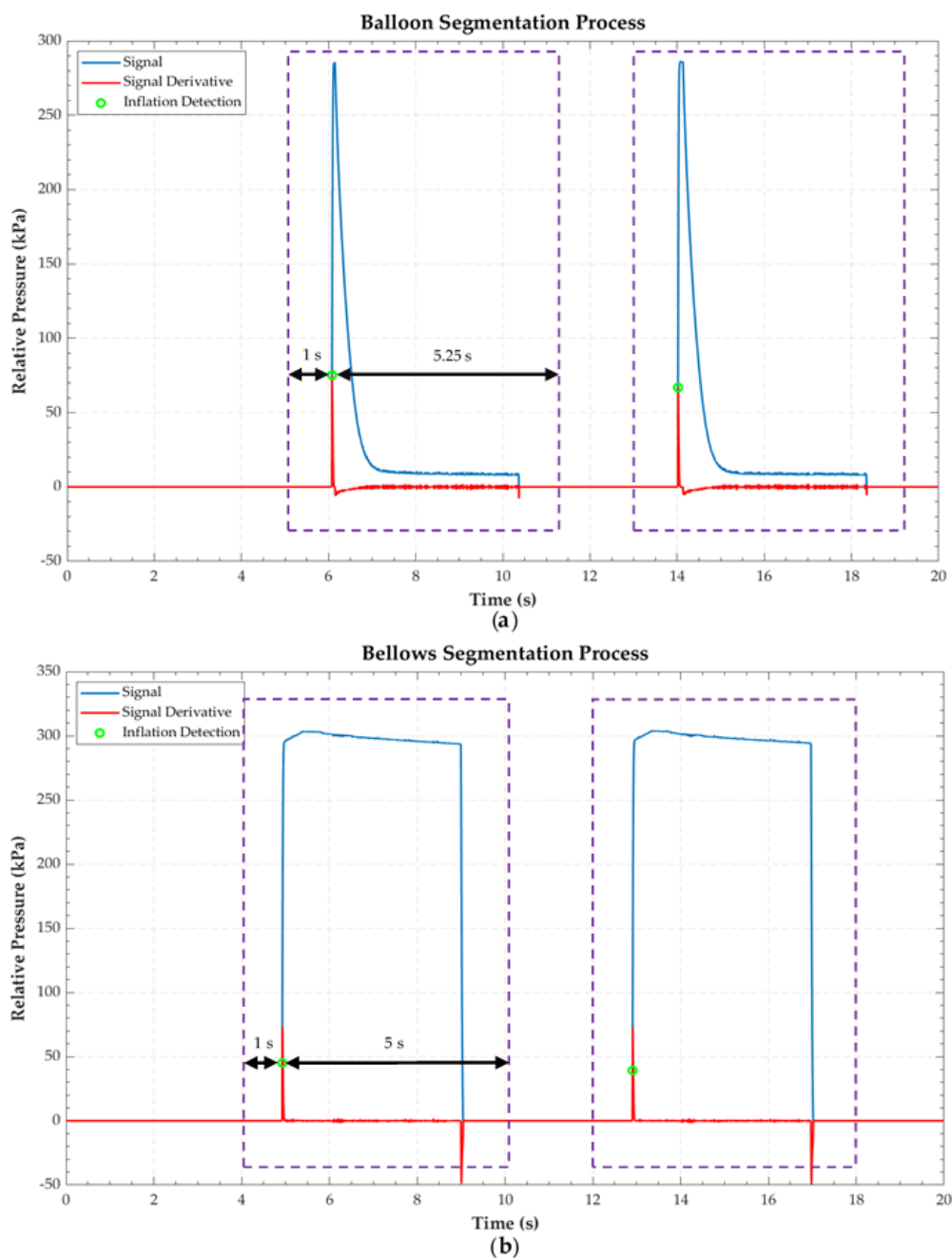into two subsets: cross-validation (80%) and testing (20%).

Figure 3. Graphical overview of the segmentation process of the two different Endoworm cavities:
(**a**) *balloon-type* and (**b**) *bellows-type*. Dashed purple linear windows correspond to the segmented signal.

Subsequently, the features of the signals were extracted. The features were identified
to reflect, quantitatively or categorically, the physical–mathematical aspects of the signals
to reflect the differences between cavities without and with leakage. In the case of the
balloon-type cavities, a total of 12 features (a_1–a_{12}) were extracted, while, for the *bellows-type*
cavities, a total of 13 features (b_1–b_{13}) were extracted.

The features of both cavity types were divided into four subgroups: a_1–a_3, a_6–a_8, and
b_1–b_8 to quantify pressure losses using different metrics; a_4 and b_4 to measure the time
difference between the detection of inflation and deflation; a_{10}, a_{11}, and b_{10}–b_{12} to indicate
the correlation of the signal concerning different standard signals of the "*Right*" or "*Leak*"
classes; a_5, a_{12}, and b_{13} to indicate categorical variables that, depending on a premise, assign
a logical value (zero or one). Specifically, a_5 indicates whether or not a deflating edge

was detected, while a_{12} and b_{13} assign a label depending on the class of the pattern with the highest correlation value between the features a_{10} and a_{11} in the *balloon and* b_{10}–b_{12} in the *bellows*, respectively. For a more extensive description of all extracted features, see Appendix A.
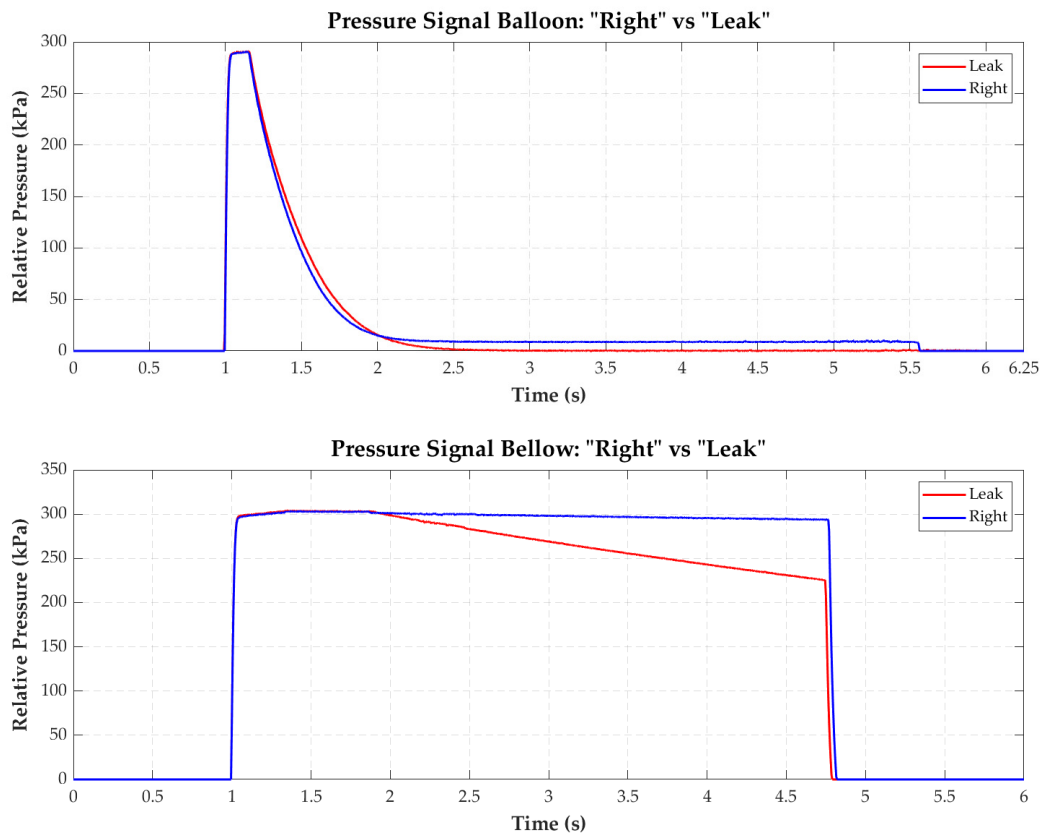

Figure 4. Representation of typical signals of the class *"Right"* (blue) and *"Leak"* (red) for the cavities *balloon* (**top**) and *bellows* (**bottom**).

In order to optimize the performance of the algorithms and make their training more efficient, the descriptors were subjected to a normalization process using the *z-score* method [22–24].

The training of the predictive models was facilitated by a prior selection of features using *filters*. To increase the discriminant potential, those variables that obtained better results with respect to an objective function were selected. A total of four *filters* were used: *Fisher's score* [25,26], *ReliefF* [27,28], *Chi-square* [29], and *MRMR* [30,31] (see Appendix B for a description of *filters*). Applying the four *filters*, four scores and ranking positions were obtained for each feature. A total score for the features was obtained by adding the ranking positions of each filter, which was lower for a better position. From this score, a total ranking was obtained that guarantees the same weight to the four *filters* when it comes to taking them into account when selecting the features. On this basis, the eight best characteristics were chosen for each type of cavity, discarding the rest.

Subsequently, the training and the validation of the predictive models were carried out. For this purpose, the *k-fold cross-validation* technique was used to make the most of the cross-validation subset, which was divided into *k-folds*. Typically, a low *k-value* means that the training subset was smaller and the validation subset was higher in percentage. This could result in a higher average prediction error (when averaging the results of the *k-folds*). In contrast, a high *k-value* would be the opposite; a higher training subset and a lower validation subset (in percent)could result in a lower average prediction error [32]. It is widely accepted to use a number of *k* between five and 20 folds [33]. A total of five folds

were used for both types of cavities. This way, five instances of training were performed with 80% of the dataset for the *cross-validation* and 20% for validation. The ratio between the validation and test set was maintained at 20%.

The "*Classification Learner*" app in "*Matlab R2022a*" was used to obtain the classification predictive models. Algorithms that require a low computational cost and memory usage were selected because the models obtained were intended to be implemented in the device's microcontroller. The following models (from the "*Classification Learner*" app) were trained: *fine, medium, and coarse tree, linear and quadratic discriminant, logistic regression, linear, quadratic and cubic SVM*, and *narrow, medium, and bilayered neural Network*, yielding a total of 12 different algorithms. For training the models, the default hyperparameter configuration of the app was maintained, and it was not modified during the whole training and validation process (*cross-validation*). It was not considered necessary to adjust the hyperparameters of each of the different algorithms.

During the cross-validation process, the wrapper technique was applied [34]. It started with the eight previously selected features, from which a reduced set that presented the best possible performance was obtained. The selected features were tested, and the test performance, together with the cross-validation, produced the final performance of each model. Due to high combinatoriality, the wrapper procedure gave rise to the training and validation of the 12 classification algorithms. Only the best-performing combinations of input features and algorithms are shown. The main metrics used for this purpose were accuracy, recall, precision, and F1-score [35]. Finally, the selected model of each type had a tradeoff among better performance, lower computational cost and memory usage, and a set of input features that were easy to compute.

3. Results

3.1. Structure of the Datasets

Tables 1 and 2 show the prevalence of the two classes, "*Right*" and "*Leak*", in the two available datasets, *balloon* and *bellows* cavities, respectively.

Table 1. Number of cases and percentage distribution of the classes for the subset of *balloon-type* data into which the initial dataset was subdivided.

Data Subset	Class	*N*	Percentage
Cross-Validation	Right	1593	84.73
	Leak	287	15.27
Test	Right	387	82.34
	Leak	83	17.66
Total	Right	1980	84.26
	Leak	370	15.74

Table 2. Number of cases and percentage distribution of the classes for the subset of *bellows-type* data into which the initial dataset was subdivided.

Data Subset	Class	*N*	Percentage
Cross-Validation	Right	637	86.20
	Leak	102	13.80
Test	Right	151	81.62
	Leak	34	18.38
Total	Right	788	85.28
	Leak	136	14.72

3.2. Preliminary Screening of Features

Box-and-whisker plots of the quantitative features of the *cross-validation* subset for each cavity type are plotted in Section 3.4 and Figure 6; categorical features were suppressed because they do not provide useful information in this type of plot. The cases were grouped according to whether they were previously classified as "*Right*" or "*Leak*".
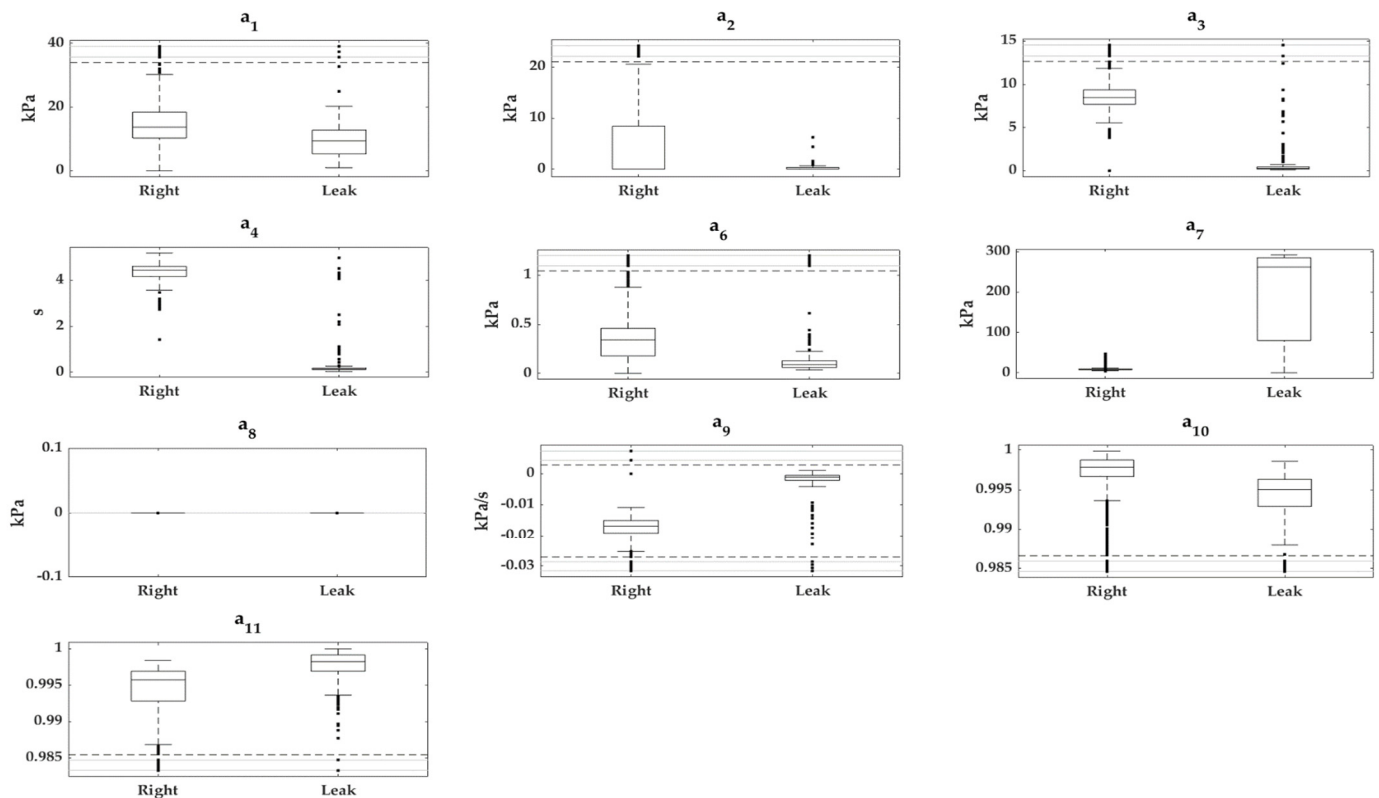

Figure 5. Box-and-whisker plots of quantitative features extracted (a_1–a_4 and a_6–a_{11}) from *balloon* signals. The "*Right*" and "*Leak*" groups are represented for each feature; a_5 and a_{12} are not presented due to their categorical nature. For additional information on the extracted features see Appendix A.

3.3. Feature Selection

Tables 3 and 4 show the results of applying the filters for the *balloon* and *bellows* features, indicating scores and rankings by features.

3.4. Classification Predictive Models

This section presents a selection of the best results obtained by training the different classification predictive models and following the assumptions specified in Section 2. Table 5 shows the main results for leak detection in the *balloon-type* cavity, where the "*Leak*" class is defined as positive, and the "*Right*" class is defined as negative.

The best-performing and simplest classification predictive model for *balloon-type* cavities was the *medium tree* algorithm with input features a_4, a_6 and a_7. Its final performance was 99.62% *accuracy*, 98.11% *recall*, 99.45% *precision*, and 98.78% *F1-score*.

Figure 7 shows the ROC curve resulting from *cross-validation* for the classification predictive model obtained with *medium tree* algorithm with input features a_4, a_6, and a_7. The ROC curve obtained for the best-performing *balloon* cavity models, shown in Table 5, is very similar to the one shown. The AUC of the best models of the *balloon* cavity was in the range of 0.97 to 0.99.
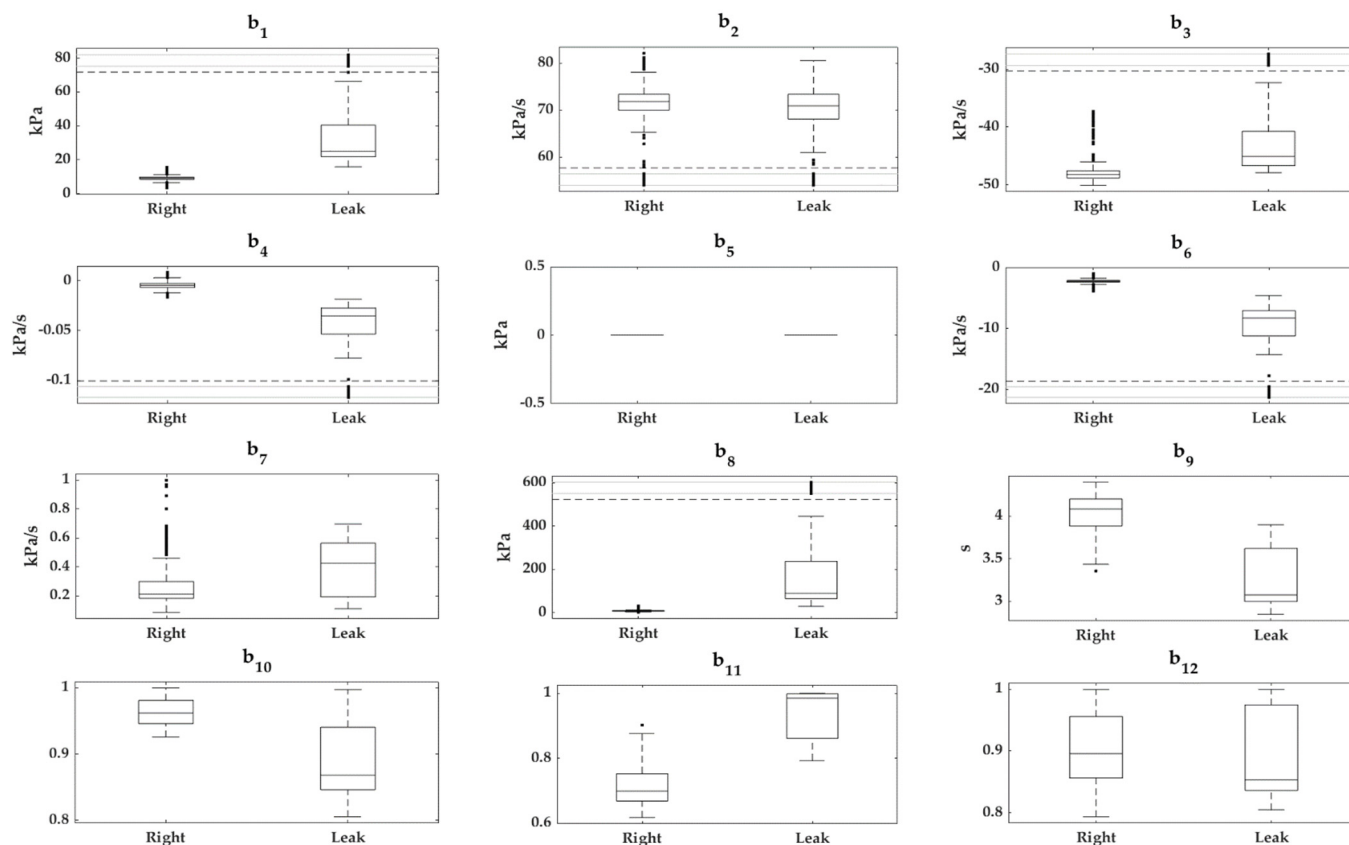
Figure 6. Box-and-whisker plots of quantitative features extracted (b_1–b_{12}) from the *bellows* signals. The "*Right*" and "*Leak*" groups are represented for each feature; b_{13} is not presented due to its categorical nature. For additional information on the extracted features see Appendix A.

Table 3. Results of applying *filters* for the features of the *balloon* signals.

Features	Scores				Ranking
	Fisher's	ReliefF	Chi^2	MRMR	Total
a_1	0.218	0.123	114.2	1.5×10^{-14}	10
a_2	0.386	0.000	160.3	2.9×10^{-14}	11
a_3	2.610	0.049	708.1	2.8×10^{-14}	5
a_4	15.824	0.073	711.9	3.9×10^{-1}	1
a_5	12.875	0.000	65,535.0	9.5×10^{-14}	2 or 3
a_6	0.002	0.054	576.7	5.9×10^{-14}	8
a_7	2.032	0.017	630.8	1.1×10^{-13}	4
a_8	0.004	-0.001	13.3	7.6×10^{-14}	12
a_9	1.644	0.050	684.8	9.8×10^{-14}	2 or 3
a_{10}	0.051	0.058	191.3	3.9×10^{-14}	9
a_{11}	0.239	0.033	347.0	3.3×10^{-1}	6
a_{12}	2.766	0.000	457.1	8.0×10^{-14}	7

Table 4. Results of applying *filters* for the features of the *bellows* signals.

| Features | Scores | | | | Ranking |
	Fisher's	ReliefF	Chi^2	MRMR	Total
b_1	0.442	0.043	278.627	0.098	5
b_2	0.017	0.000	2.338	0.002	12
b_3	0.433	0.010	99.692	0.099	9
b_4	0.570	0.034	278.873	0.199	2 or 3
b_5	0.000	0.000	0.000	0.000	13
b_6	0.579	0.042	278.873	0.131	2 or 3
b_7	0.283	−0.017	38.996	0.061	11
b_8	0.127	0.004	278.566	0.387	6 or 7
b_9	3.009	0.009	218.732	0.061	6 or 7
b_{10}	1.380	0.019	211.381	0.332	4
b_{11}	4.779	0.019	266.701	0.259	1
b_{12}	0.010	0.007	40.772	0.008	10
b_{13}	1.751	0.000	114.625	0.128	8

Table 5. The seven models that obtained the best results for leak detection in *balloon-type* cavities. Results are expressed as percentages.

| Algoritm | Features | 5-Fold Cross-Validation | | | | Test | | | | Total | | | |
		Acc.	Recall	Pre.	F1-Sc.	Acc.	Recall	Pre.	F1-Sc.	Acc.	Recall	Pre.	F1-Sc.
Medium tree	a_4, a_6, a_7	99.63	98.26	99.30	98.77	99.57	97.59	100	98.78	99.62	98.11	99.45	98.78
	$a_3{-}a_7, a_9, a_{11}, a_{12}$	99.63	98.26	99.30	98.77	99.57	97.59	100	98.78	99.62	98.11	99.45	98.78
	a_4, a_6	99.36	97.91	97.91	97.91	99.79	98.80	100	99.39	99.45	98.11	98.37	98.24
Bilayered neural network	$a_3{-}a_7, a_9, a_{11}, a_{12}$	99.41	97.21	98.94	98.07	99.57	97.59	100	98.78	99.45	97.30	99.17	98.23
Narrow neural network	a_4, a_6, a_7, a_9	99.36	97.21	98.59	97.89	99.79	98.80	100	99.39	99.45	97.57	98.90	98.23
Quadratic SVM	$a_4{-}a_7$	99.36	96.17	99.64	97.87	99.57	97.59	100	98.78	99.40	96.49	99.72	98.08
Medium neural network	$a_4{-}a_7$	99.36	98.26	97.58	97.92	99.36	96.39	100	98.16	99.36	97.84	98.10	97.97

Representative validation ROC curve for the balloon cavity
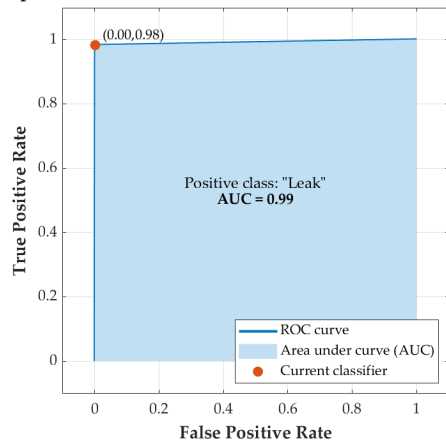
Positive class: "Leak"
AUC = 0.99

Figure 7. Representative validation ROC curve obtained for the *balloon* cavity model with *median tree* algorithm with input features a_4, a_6, and a_7.

Figure 8 shows the scatter plots of the three input features to this predictive model. It also indicates the cases that were hit and miss in the model's classification of the entire dataset.

On the other hand, the models of the *bellows* cavities obtained perfect results, with a correct classification for all the cases presented (100% *accuracy*). The best-performing algorithms were *logistic regression*, *linear SVM*, *neural networks*, and *coarse tree*, when features b_1, b_4, and b_6 were part of the input set to the models.

Figure 9 shows the ROC curve resulting from *cross-validation* for the classification predictive model obtained with the *logistic regression* algorithm with input features b_1, b_4, and b_6. The ROC curve obtained for the best-performing *bellows* cavity models is identical to the one shown. The AUC was 1.00 for all models.

Figure 10 shows the scatter plots of the three input features to this classification predictive model. It also indicates the cases that were hit and miss in the model's classification of the entire dataset.
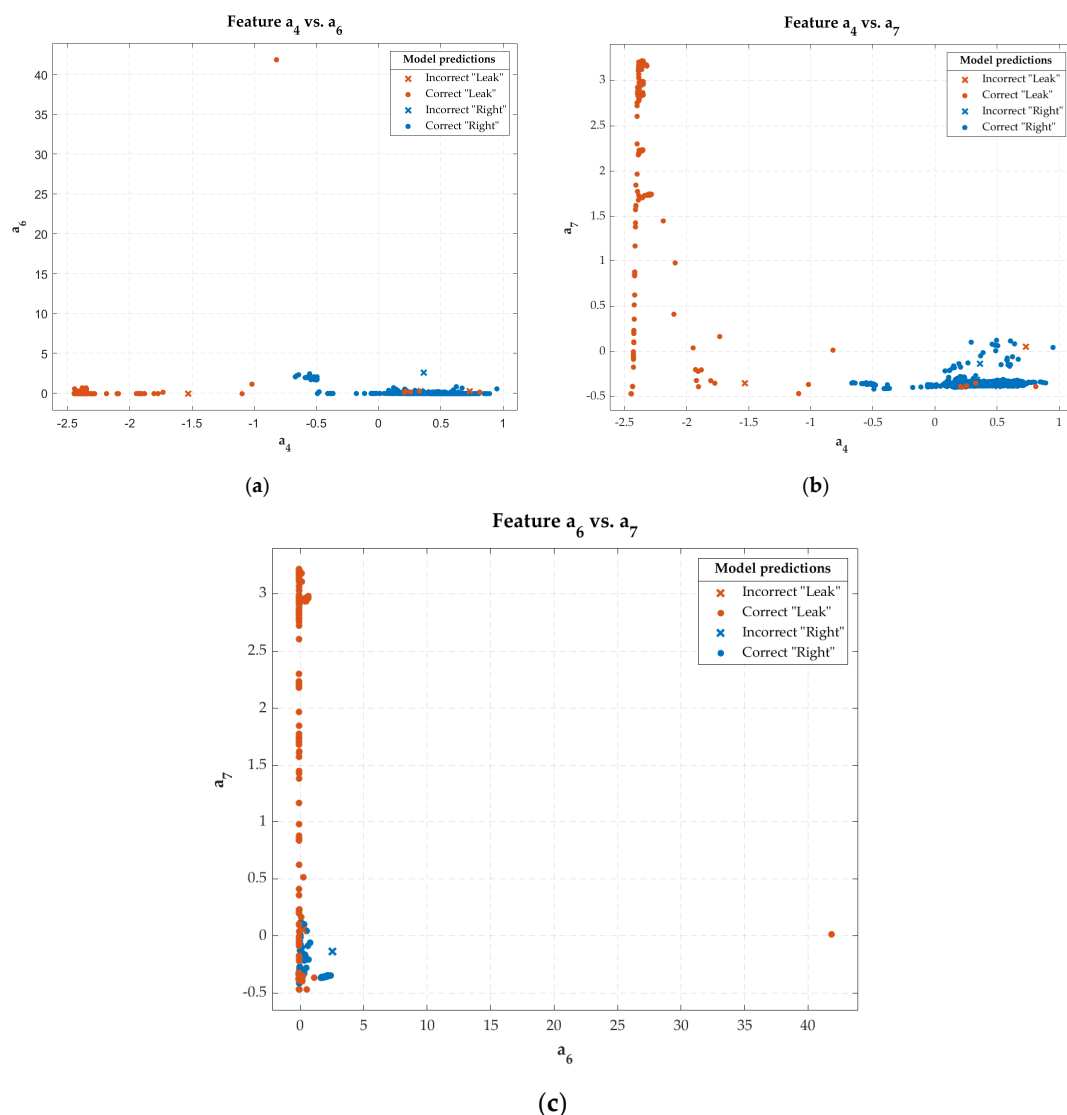

Figure 8. Scatter plots with the classification results of the predictive *medium tree* model with the input features a_4, a_6, and a_7: scatter plot of feature a_4 vs. a_6 (**a**); scatter plot of feature a_4 vs. a_7 (**b**); scatter plot of feature a_6 vs. a_7 (**c**).
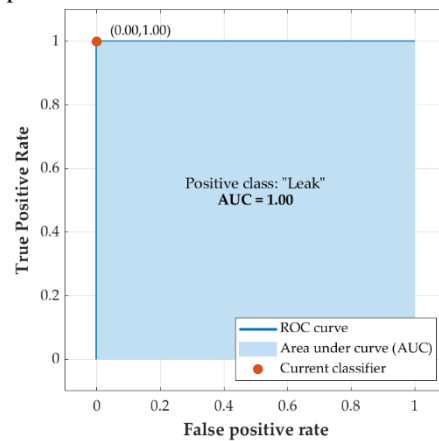
Figure 9. Representative validation ROC curve obtained for the *bellows* cavity model with *logistic regression* with input features b_1, b_4, and b_6.
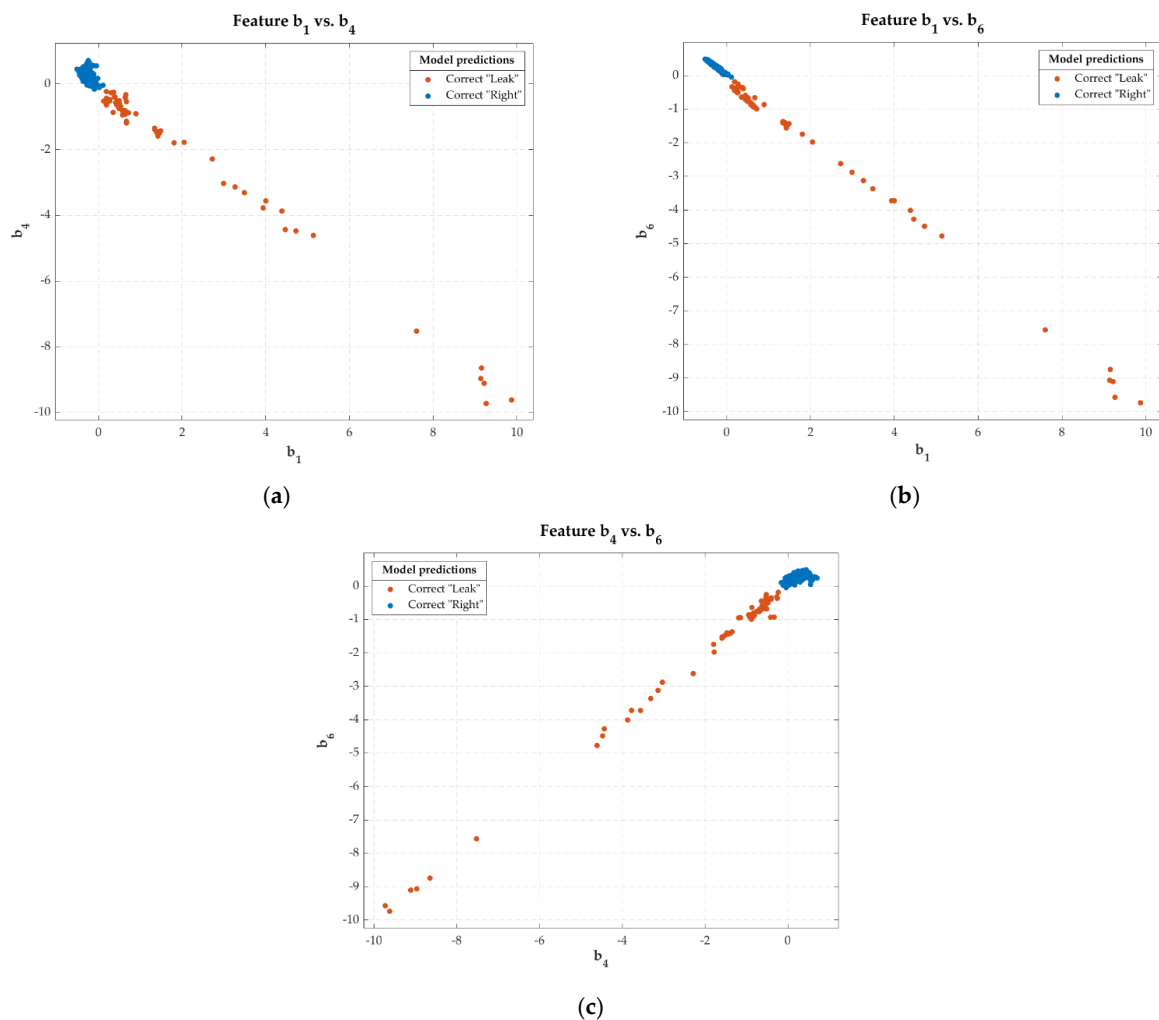

(**a**)

(**b**)

(**c**)

Figure 10. Scatter plots with the classification results of the predictive *logistic regression* model with the input features b_1, b_4, and b_6: scatter plot of feature b_1 vs. b_4 (**a**); scatter plot of feature b_1 vs. b_6 (**b**); scatter plot of feature b_4 vs. b_6 (**c**).

4. Discussion

The results showed an imbalance in the two datasets (*balloon* and *bellows)*, being around 85% for the "*Right*" class versus 15% for "*Leak*" for both (Tables 1 and 2). This fact must be considered to evaluate the performance of the predictive classification. FN (false negative) mean that the device does not detect air leakage, resulting in unnecessary air injection into the patient's small bowel, while FP (false positive) is the opposite case, resulting in unnecessary interruption of the scan due to false detections of leakage, which is also undesirable for the device's proper functioning. In the issue at hand, it is preferable to give priority to FNs over FPs, as this prioritizes patient safety over the interruption of the scan. Therefore, the model with the best F1-score result and minimum number of FN was prioritized, i.e., the one with the highest *recall*.

In the exploration of the cavity features, it was detected that a_8 and b_5 had a value of 0 kPa in all the samples (Section 3.4 and Figure 6). This did not provide any discriminant capacity to our predictive models; hence, it can be anticipated that they will obtain very low scores when applying the *filters*. These two features belong to the subgroup of pressure parameters; more specifically, both indicate the pressure 0.5 s before cavity inflation occurs.

Section 3.4 and Figure 6 show that the features a_1, a_2, a_6, a_7, a_{10}, a_{11}, b_2, b_3, b_7, and b_9–b_{12} overlapped with the boxes and/or whiskers of the two classes, indicating that, on their own, they did not appear to have high discriminatory power between classes. In contrast, the features a_3, a_4, a_9, b_1, b_4, b_6, and b_8 did not have overlapping boxes or whiskers between classes, but presented anomalous data that overlapped with the boxes and whiskers or anomalous data from the opposite class. The latter group of features had a high interclass discriminant potential on their own.

Following the preliminary analysis of the features, the *filters* were applied; the results are shown in Tables 3 and 4 for the *balloon* and the *bellows*, respectively. Calculating the total ranking, it was guaranteed that the four *filters* provided the exact weighting. On this basis, the eight best characteristics of each cavity type were selected, and the rest were discarded. Thus, a_1, a_2, a_8, and a_{12} were discarded for *balloon-type* cavity models, and b_2, b_3, b_5, b_7, and b_{12} were discarded for bellows.

As mentioned in Section 3, the AUC ranges obtained in the training of the best models were between 0.97 and 0.99 for the *balloon* cavity models and 1.00 for the *bellows* cavity models. This indicates that the performances obtained by the models were excellent. Additionally, in the representative ROC curves of the models (see Figures 7 and 9), it can be seen that the training of the models converged with few iterations, and that they were very close to the ideal ROC.

Regarding the models obtained, it can be said that the algorithms that performed best from highest to lowest performance for *balloon* cavities were *medium tree, neural networks*, and *quadratic discriminant*. Furthermore, most of the interclass discriminating power, regardless of the algorithm, was concentrated in the features a_4, a_6, and a_7 (see Table 5). It was observed that the combination of a_4 with a_6 and a_7 gave high discriminatory power, as the cases of different classes tend to be grouped, except for some cases of "*Leak*" that were embedded within the cluster of cases of the "*Right*" class. However, a_4 and a_7 did not have such a clear separation of cases and tended to intermingle to a greater extent (see Figure 8).

On the other hand, the predictive models of the *bellows* cavities obtained perfect results, with a correct classification for all the cases presented. The algorithms with the best performance for this case were *logistic regression, linear SVM, neural networks*, and *coarse tree.* These excellent results could be explained by the absolute discriminant power of the b_6 feature, which was capable of 100% classification accuracy in the vast majority of algorithms. Excellent predictive power was obtained if this was combined with the b_4 feature (see Figure 10).

For the detection of air leaks in *balloon* cavities, we chose the classification predictive model obtained with the *medium tree* algorithm and the input features a_4, a_6, and a_7. It had a final performance of 99.62% *accuracy*, 98.11% *recall*, 99.45% precision, and 98.78% *F1-score*. The model achieved the best result for the *F1-score/recall* ratio. Moreover, this model was

chosen due to its low computational cost and memory usage, and the input features were simple to calculate.

Tree-type algorithms are implemented straightforwardly by nesting *"if–else"* statements that compare a given threshold of input features to the model (establishment of decision boundaries).

It had as input features a_4, a_6, and a_7, which were simple to calculate and required a low computational cost compared to other features. The feature a_4 was the time in seconds from detecting balloon cavity inflation to detecting deflation. The feature a_6 was the variability of the cavity pressure in kPa in the time span from when the pressure signal stabilized after the inflation transient to the detection of cavity deflation. The feature a_7 was the average pressure in the time range between 100 and 50 ms before detecting deflation of the signal.

The only features that were computationally and memory-intensive were those that belonged to or were directly related to the subgroup of features that showed the correlation of the signal to different standard signals of the *"Right"* or *"Leak"* classes (a_{10}–a_{12} and b_{10}–b_{13}).

To detect air leaks in the *bellows* cavities, we chose the classification predictive model obtained with the *logistic regression* algorithm with the input features b_1, b_4, and b_6, achieving a performance of 100% in all metrics. In addition to presenting unbeatable results, it was selected because a predictive *logistic regression* model is easy to implement on a microcontroller and requires a low computational cost. Moreover, the input features it used were easy to compute. The feature b_1 was the difference (in kPa) between the pressure in the *bellow* cavity when inflation was detected and the pressure just before deflation was detected. The feature b_4 was the average of the derivative of the pressure signal (kPa/s) at the instant between the detection of cavity inflation and the instant just before detection of cavity inflation. In contrast, feature b_6 estimated the pressure slope (kPa/s) calculated as feature b_1 divided by the time elapsed between the detection of cavity inflation and deflation.

5. Conclusions

This article aimed to obtain two classification predictive models, one for each type of cavity (*balloon* and *bellows*), that detect the presence of air leaks from the cavity and that could be implemented within the Endoworm control device. The models served two purposes: safety for patients and effective functioning. The most important was patient safety against possible problems due to excess air insufflation in the small intestine after an air leak. The second was to provide the device with a mechanism to ensure its effective functioning. If any cavities leak air, the translation system will not effectively perform its function (i.e., to fix and retract the intestine over the endoscope advancing through the digestive tract).

In order to achieve this objective, a series of features were extracted after digital processing of the pressure signals. Features were analyzed and selected by filtering. Later, *fivefold cross-validation* with *wrappers* for the training and validation of different supervised classification predictive models was applied. Finally, these models were tested with a subsequent selection of the best set of features and algorithm that obtained the best results.

Following this procedure, it was concluded that the best model for air leakage detection in *balloon* cavities was crafted from the *medium tree* algorithm and the input features a_4, a_6, and a_7. The best model for the bellows cavities was composed of the *logistic regression* algorithm and the input features b_1, b_4, and b_6.

With regard to the features extracted from the signals of the *balloon* cavities, it can be concluded that the best discriminating results were obtained by combining that which measures the time the cavity remained inflated (a_4) with two parameters that measured the pressure and variability of the pressure once the cavity inflation pressure stabilized (a_7 and a_6, respectively). On the other hand, the features extracted from the signals of the *bellows* cavities were those that measured the pressure loss while the cavity remained inflated, either incrementally (b_1) or through the derivative or approximations thereof (b_4 and b_5, respectively).

The algorithm and features used in both models required relatively low computational costs, allowing them to be run on the microcontroller of the Endoworm control device

without problems. This will allow detecting, practically in real time, the presence or absence of air leaks in the medical device during small bowel examinations in patients.

The results obtained in this article are encouraging and meet the proposed objectives, as models with a high degree of reliability were obtained. In the future, the models obtained in this article should be tested and validated with experimental data on animal models (pigs are generally used) and subsequently with clinical trials on patients.

Author Contributions: Conceptualization, C.S.-D. and A.V.; methodology, R.Z.-M.; software, R.Z.-M.; validation, V.P.-B., C.S.-D. and R.Z.-M.; formal analysis, C.S.-D.; investigation, V.P.-B., A.V. and R.Z.-M.; resources, V.P.-B. and C.S.-D.; data curation, C.S.-D.; writing—original draft preparation, R.Z.-M.; writing—review and editing, V.P.-B., A.V. and C.S.-D.; visualization, A.S.; supervision, C.S.-D.; funding acquisition, V.P.-B., A.V. and C.S.-D. All authors read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the dataset is continually being expanded with new in vitro tests. In addition, it is expected that the data set will be completed in the future with in vivo tests in animals and later in humans, as mentioned in the conclusions of the paper. For this reason, if you would like an updated version of the dataset, please contact the authors directly.

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

Appendix A

Bellow: the description of the features that were extracted from the pressure signals of the *balloon* (a_i) and *bellows* (b_i) type cavities are shown below.

a_1: Pressure drop occurring in the signal's stationary regime while the cavity remains inflated. It is calculated as the difference between the pressure value 1.5 and 3 s after the detection of cavity inflation (kPa).

a_2: Pressure value close to the deflation of the signal. This is the pressure value 3 s after cavity inflation is detected (kPa).

a_3: Pressure in the stationary regime when the cavity is inflated. It is calculated as the average of sample pressures between 1.5 and 3 s after the detection of cavity inflation (kPa).

a_4: Time between detection of inflation and deflation of the cavity. It is calculated as the difference between the instant of time at which deflation is detected with inflation (s).

a_5: Categorical feature indicating whether cavity deflation is detected. Its value is '*1*' if a negative peak in the derivative of the pressure signal is detected in the time interval between 1.5 and 4 s after the detection of cavity inflation; otherwise, its value is '*0*' (dimensionless).

a_6: Variability of pressure in the stationary regime when the cavity remains inflated. It is calculated as the statistical variability of the pressure values between the time instants 1.5 and 3 s after the detection of cavity inflation (kPa).

a_7: Average pressure immediately before to the detection of cavity deflation. It is calculated as the average pressure between 100 and 50 ms before cavity deflation is detected (kPa).

a_8: Average pressure before the cavity is inflated. It is calculated as the average pressure of the values between 550 and 450 ms before cavity inflation is detected (kPa).

a_9: Average cavity deflation slope in the stationary regime when the cavity is inflated. It is calculated as the average of the derivative of the pressure values between 1.5 s after cavity inflation and one sample before signal deflation is detected. In the case where cavity

deflation is not detected (feature a_5), it is calculated as the average of values between 1.5 and 3 s after the detection of signal inflation (kPa/s).

a_{10}: Maximum correlation of the segmented pressure signal concerning four standard signals of the class "*Right*". It is calculated by obtaining the correlation of the four standard signals concerning the segmented signal, assigning to the value of the feature the highest value of the four correlations obtained (dimensionless).

a_{11}: Maximum correlation of the segmented pressure signal concerning four standard signals of the class "*Leak*". It is calculated using the same method as a_{10} (dimensionless).

a_{12}: Categorical feature indicating whether the signal is of the "*Right*" or "*Leak*" class on the basis of the correlation value of the signal concerning the standard signals of the two classes. It is assigned a value of '*0*' if a_{10} is greater than a_{11} and '*1*' otherwise (dimensionless).

b_1: Pressure loss that occurs while the cavity is inflated. It is calculated as the difference between the maximum pressure from the time cavity inflation is detected until 1 s later and the minimum pressure from the time cavity deflation is detected until 1 s earlier (kPa).

b_2: Maximum peak of the derivative of the segmented pressure signal, corresponding to cavity inflation. It is calculated as the maximum value of the derivative of the signal (kPa/s).

b_3: Minimum peak of the derivative of the segmented pressure signal, corresponding to cavity deflation. It is calculated as the minimum value of the derivative of the signal (kPa/s).

b_4: Average slope of the pressure loss that occurs while the cavity remains inflated. It is calculated as the average of the derivative values between the maximum peak (inflated) and the minimum peak (deflated) of the derivative of the pressure signal (kPa/s).

b_5: Average pressure before to cavity inflation. It is calculated as the average pressure between 200 and 250 ms before detection of cavity inflation (kPa).

b_6: Estimation of the pressure loss slope while the cavity remains inflated. It is calculated as the feature b_1 divided by the time difference between the detection of deflation and inflation of the cavity (kPa/s).

b_7: Variability of the pressure loss slope while the cavity remains inflated. It is calculated as the statistical variability of the values of the pressure derivative between the maximum and minimum peak of the derivative (kPa/s).

b_8: Variability of pressure while the cavity remains inflated. It is calculated as the statistical variability of the pressure values between inflation and deflation detection (kPa).

b_9: Time the cavity remains inflated. It is calculated as the difference between the time instant at which deflation and inflation are detected (s).

b_{10}: Maximum correlation of the segmented pressure signal concerning three standard signals of the class "*Right*". It is calculated by obtaining the correlation of the three standard signals with respect to the segmented signal, assigning to the value of the feature the highest value of the four correlations obtained (dimensionless).

b_{11}: Maximum correlation of the segmented pressure signal concerning two standard signals of the class "*Leak*". It is calculated using the same method as b_{10} (dimensionless).

b_{12}: Maximum correlation of the segmented pressure signal concerning two standard signals of the class "*Leak*". It is calculated using the same method as b_{10}, but the standard signals are different from those used in b_{11} (dimensionless).

b_{13}: Categorical feature indicating whether the signal is of the "*Right*" or "*Leak*" class on the basis of the correlation value of the signal concerning the standard signals of the two classes. It is assigned a value of '*0*' if b_{10} is greater than b_{11} or b_{12} and '*1*' otherwise (dimensionless).

Appendix B

The formulation of the four types of filters used for the selection of the features is shown below.

1. ***Fisher's score*** was calculated according to Equation (A1).

$$F(k) = \frac{\sum_{i=1}^{c} n_i \left(\mu_k^i - \mu_k\right)^2}{\sum_{i=1}^{c} n_i \left(\sigma_k^i\right)^2} , \tag{A1}$$

where n_i is the number of the i-th class, μ_k^i and σ_k^i are the mean and variance of the k-th feature in the i-th class, respectively, and μ_k is the mean of the k-th feature in all classes [26].

2. ***ReliefF score*** was calculated using the default algorithm in *Matlab 2022a* using the "*relieff*" function. This algorithm works best for estimating feature importance for distance-based supervised models that use pairwise distances between observations to predict the response. The theoretical development of the calculations made by the algorithm to obtain the feature scores was described in [27].

3. ***Chi-square score*** was calculated using the default algorithm in *Matlab 2022a* using the "*fscchi2*" function. This filter examines whether each predictor variable is independent of a response variable using individual *chi-square tests*, ranking the features using the p-values of the *chi-square* test statistics. The theoretical development of how the algorithm works to obtain the scores and the ranking was shown in [29].

4. ***MRMR score*** was calculated using the default algorithm in *Matlab 2022a* using the "*fscmrmr*" function. The MRMR algorithm is responsible for finding an optimal set of features that are mutually and maximally dissimilar and that can effectively represent the response variable. The algorithm's goal is to minimize the redundancy of a set of features and maximize the relevance of a set of features to the response variable. The algorithm quantifies redundancy and relevance using the mutual information of the variables (by pairs of features) and mutual information between a feature and the response. The theoretical and mathematical development of the algorithm was shown in [31].

References

1. Tsujikawa, T.; Saitoh, Y.; Andoh, A.; Imaeda, H.; Hata, K.; Minematsu, H.; Senoh, K.; Hayafuji, K.; Ogawa, A.; Nakahara, T.; et al. Clinical impact of novel single balloon enteroscopy. *J. Gastroenterol. Hepatol.* **2007**, 22, A226.
2. Yamamoto, H.; Sekine, Y.; Sato, Y.; Higashizawa, T.; Miyata, T.; Iino, S.; Ido, K.; Sugano, K. Total enteroscopy with a nonsurgical steerable double-balloon method. *Gastrointest. Endosc.* **2001**, 53, 216–220. [CrossRef] [PubMed]
3. Sunada, K.; Yamamoto, H. Double-balloon endoscopy: Past, present, and future. *J. Gastroenterol.* **2019**, 44, 1–12. [CrossRef]
4. Akerman, P.A.; Haniff, M. Spiral enteroscopy: Prime time or for the happy few? *Best Pract. Res. Clin. Gastroenterol.* **2012**, 26, 293–301. [CrossRef]
5. Lenz, P.; Domagk, D. Double- vs. single-balloon vs. spiral enteroscopy. *Best Pract. Res. Clin. Gastroenterol.* **2012**, 26, 303–313. [CrossRef] [PubMed]
6. Wadhwa, V.; Sethi, S.; Tewani, S.; Garg, S.K.; Pleskow, D.K.; Chuttani, R.; Berzin, T.M.; Sethi, N.; Sawhney, M.S. A meta-analysis on efficacy and safety: Single-balloon vs. double-balloon enteroscopy. *Gastroenterol. Rep.* **2015**, 3, 148–155. [CrossRef] [PubMed]
7. May, A.; Manner, H.; Aschmoneit, I.; Ell, C. Prospective, cross-over, single-center trial comparing oral double-balloon enteroscopy and oral spiral enteroscopy in patients with suspected small-bowel vascular malformations. *Endoscopy* **2011**, 43, 477–483. [CrossRef]
8. Kim, T.J.; Kim, E.R.; Chang, D.K.; Kim, Y.H.; Hong, S.N. Comparison of the efficacy and safety of single-versus double-balloon enteroscopy performed by endoscopist experts in single-balloon enteroscopy: A single-center experience and meta-analysis. *Gut Liver* **2017**, 11, 520–527. [CrossRef]
9. Nehme, F.; Goyal, H.; Perisetti, A.; Tharian, B.; Sharma, N.; Tham, T.C.; Chhabra, R. The Evolution of Device-Assisted Enteroscopy: From Sonde Enteroscopy to Motorized Spiral Enteroscopy. *Front. Med.* **2021**, 8, 792668. [CrossRef]
10. Qi, Q.; Teng, Y.; Li, X. Design and characteristic study of a pneumatically actuated earthworm-like soft robot. In Proceedings of the 2015 International Conference on Fluid Power and Mechatronics (FPM), Harbin, China, 5–7 August 2015; pp. 435–439. [CrossRef]
11. Dewapura, J.I.; Hemachandra, P.S.; Dananjaya, T.; Awantha, W.V.I.; Wanasinghe, A.T.; Kulasekera, A.L.; Chathuranga, D.S.; Dassanayake, V.P.C. Design and development of a novel bio-inspired worm-type soft robot for in-pipe locomotion. In Proceedings of the 2020 20th International Conference on Control, Automation and Systems (ICCAS), Busan, Korea, 13–16 October 2020; pp. 586–591. [CrossRef]
12. Calisti, M.; Picardi, G.; Laschi, C. Fundamentals of soft robot locomotion. *J. R. Soc. Interface* **2017**, 14, 20170101. [CrossRef]
13. Tang, Z.; Lu, J.; Wang, Z.; Ma, G.; Chen, W.; Feng, H. Development of a New Multi-cavity Pneumatic-driven Earthworm-like Soft Robot. *Robotica* **2020**, 38, 2290–2304. [CrossRef]
14. Tobella, J.; Pons-Beltrán, V.; Santonja, A.; Sánchez, C.; Campillo-Fernández, A.J.; Vidaurre, A. Analysis of the 'Endoworm' prototype's ability to grip the bowel in in vitro and ex vivo models. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **2020**, 234, 468–477. [CrossRef] [PubMed]
15. Sánchez-Diaz, C.; Senent-Cardona, E.; Pons-Beltran, V.; Santonja-Gimeno, A.; Vidaurre, A. Endoworm: A new semi-autonomous enteroscopy device. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **2018**, 232, 1137–1143. [CrossRef] [PubMed]

16. Gerson, L.B.; Flodin, J.T.; Miyabayashi, K. Balloon-assisted enteroscopy: Technology and troubleshooting. *Gastrointest. Endosc.* **2008**, *68*, 1158–1167. [CrossRef] [PubMed]

17. Waleed, D.; Mustafa, S.H.; Mukhopadhyay, S.; Abdel-Hafez, M.F.; Jaradat, M.A.K.; Dias, K.R.; Arif, F.; Ahmed, J.I. An In-Pipe Leak Detection Robot with a Neural-Network-Based Leak Verification System. *IEEE Sens. J.* **2019**, *19*, 1153–1165. [CrossRef]

18. da Cruz, R.P.; da Silva, F.V.; Fileti, A.M.F. Machine learning and acoustic method applied to leak detection and location in low-pressure gas pipelines. *Clean Technol. Environ. Policy* **2020**, *22*, 627–638. [CrossRef]

19. Kosturkov, R.; Nachev, V.; Titova, T. Diagnosis of Pneumatic Systems on Basis of Time Series and Generalized Featurefor Comparison with Standards for Normal Working Condition. *TEM J.* **2021**, *10*, 183–191. [CrossRef]

20. Geiger, G. State-of-the-art in leak detection and localization. *Oil Gas-European Mag.* **2006**, *122*, 193–198.

21. Wu, C.-T.; Li, G.-H.; Huang, C.-T.; Cheng, Y.-C.; Chen, C.-H.; Chien, J.-Y.; Kuo, P.-H.; Kuo, L.-C.; Lai, F. Acute exacerbation of a chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: Development and cohort study. *JMIR mHealth uHealth* **2021**, *9*, e22591. [CrossRef]

22. Lee, E.A.; Seshia, S.A. *Introduction to Embedded Systems. A Cyber-Physical Systems Approach*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2017; Volume 215.

23. Jain, A.; Nandakumar, K.; Ross, A. Score normalization in multimodal biometric systems. *Pattern Recognit* **2005**, *38*, 2270–2285. [CrossRef]

24. Haga, A.; Takahashi, W.; Aoki, S.; Nawa, K.; Yamashita, H.; Abe, O.; Nakagawa, K. Standardization of imaging features for radiomics analysis. *J. Med. Investig.* **2019**, *66*, 35–37. [CrossRef] [PubMed]

25. Mahanta, M.S.; Plataniotis, K.N. Ranking 2DLDA features based on Fisher Discriminance. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 8307–8311. [CrossRef]

26. Gan, M.; Zhang, L. Iteratively local fisher score for feature selection. *Appl. Intell.* **2021**, *51*, 6167–6181. [CrossRef]

27. Robnik-Sikonja, M.; Kononenko, I. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Mach. Learn.* **2003**, *53*, 23–69. [CrossRef]

28. Wang, Z.; Zhang, Y.; Chen, Z.; Yang, H.; Sun, Y.; Kang, J.; Yang, Y.; Liang, X. Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image. *Int. Geosci. Remote Sens. Symp.* **2016**, *2016*, 755–758. [CrossRef]

29. Liu, H.; Setiono, R. Chi2: Feature selection and discretization of numeric attributes. In Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence, Herndon, VA, USA, 5–8 November 1995; pp. 388–391. [CrossRef]

30. Sun, Y.; Ma, L.; Qin, N.; Zhang, M.; Lv, Q. Analog filter circuits feature selection using MRMR and SVM. In Proceedings of the 2014 14th International Conference on Control, Automation and Systems (ICCAS 2014), Gyeonggi-do, Korea, 22–25 October 2014; pp. 1543–1547. [CrossRef]

31. Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. In Proceedings of the 2003 IEEE Bioinformatics Conference, CSB2003, Stanford, CA, USA, 11–14 August 2003; pp. 523–528. [CrossRef]

32. Rodríguez, J.D.; Pérez, A.; Lozano, J.A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [CrossRef]

33. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining. Practical Machine Learning Tools and Techniques*, 3rd ed.; Elsevier: Burlington, MA, USA, 2008.

34. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]

35. Kim, M.J. Building a cardiovascular disease prediction model for smartwatch users using machine learning: Based on the Korea national health and nutrition examination survey. *Biosensors* **2021**, *11*, 228. [CrossRef]