

## Article

# Deep Learning Algorithm for Heart Valve Diseases Assisted Diagnosis

Santiago Isaac Flores-Alonso <sup>1,†</sup> , Blanca Tovar-Corona <sup>2,†</sup>  and René Luna-García <sup>1,\*,†</sup> <sup>1</sup> CIC-IPN, Ciudad de México 07738, Mexico; sfloresa2010@alumno.ipn.mx<sup>2</sup> UPIITA-IPN, Ciudad de México 07340, Mexico; bltovar@ipn.mx

\* Correspondence: rlunag@ipn.mx

† These authors contributed equally to this work.

**Abstract:** Heart sounds are mainly the expressions of the opening and closing of the heart valves. Some sounds are produced by the interruption of laminar blood flow as it turns into turbulent flow, which is explained by abnormal functioning of the valves. The analysis of the phonocardiographic signals has made it possible to indicate that the normal and pathological records differ from each other concerning both temporal and spectral features. The present work describes the design and implementation based on deep neural networks and deep learning for the binary and multiclass classification of four common valvular pathologies and normal heart sounds. For feature extraction, three different techniques were considered: Discrete Wavelet Transform, Continuous Wavelet Transform and Mel Frequency Cepstral Coefficients. The performance of both approaches reached F1 scores higher than 98% and specificities in the “Normal” class of up to 99%, which considers the cases that can be misclassified as normal. These results place the present work as a highly competitive proposal for the generation of systems for assisted diagnosis.

**Keywords:** deep learning; CWT; deep neural networks; DWT; MFCCs; phonocardiography; valvular disease

**Citation:** Flores-Alonso, S.I.;

Tovar-Corona, B.; Luna-García, R.

Deep Learning Algorithm for Heart  
Valve Diseases Assisted Diagnosis.*Appl. Sci.* **2022**, *12*, 3780. <https://doi.org/10.3390/app12083780>Academic Editors: Keun Ho Ryu and  
Nipon Theera-Umporn

Received: 25 February 2022

Accepted: 6 April 2022

Published: 8 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

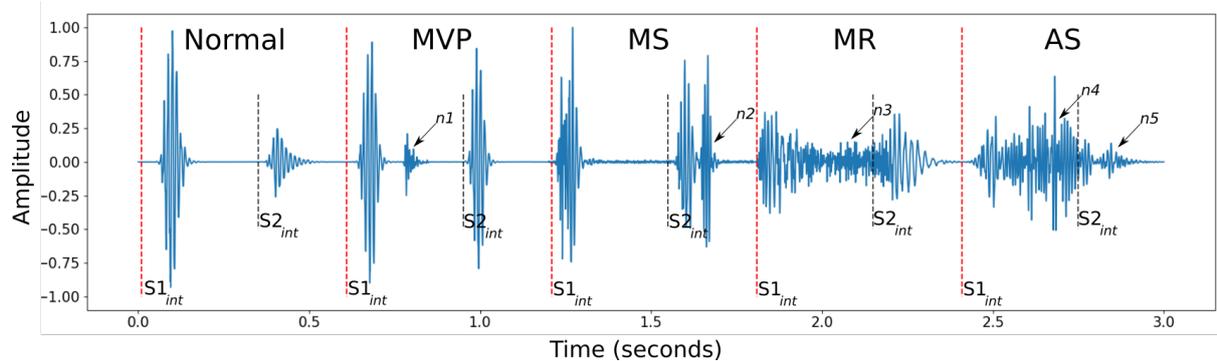
## 1. Introduction

Cardiovascular diseases occupy first place among the causes of death around the world, according to the World Health Organization (WHO) [1]. Heart valve diseases (HVD) are also found in these figures, where moderate or severe valvular abnormalities are notably common in the adult population and increase their presence as the individuals age [2].

To examine the condition of the heart valves, the most common medical practice is auscultation, which consists of listening to acoustic characteristics directly via the patient's chest wall using a stethoscope. These heart sounds could be interpreted as the acoustic expression of the opening and closing of the four heart valves—tricuspid, mitral, pulmonary and aortic—where the muscular contraction that drives the blood from one cavity to another generates a high acceleration and delay of the blood flow, causing a pressure difference [3,4]. Its normal physiological functioning is always unidirectional, which allows the correct circulation of blood through the cardiovascular circuit. However, some sounds are produced by the interruption of laminar blood flow by turning into turbulent flow, which is explained by abnormal and pathological functioning of the heart valves.

The cardiac cycle is composed of two phases: the systole, during which the ventricles contract and drive blood to the blood vessels, and the diastole, in which the ventricles are filled. The systole begins with the closure of the mitral and tricuspid valves, producing the first heart sound or S1, while the diastole starts with the closure of the aortic and pulmonic valves, producing the second heart sound or S2. In addition to S1 and S2, extra sounds could be present during the cardiac cycle, which may indicate an abnormality [5,6]. The duration of the noise varies depending on the valvular abnormality [7]. Figure 1 illustrates

the approximate duration of the heart sounds present in a normal cardiac cycle and heart noises in different valvular abnormalities.



**Figure 1.** Heart sounds (S1, S2) and noises ( $nX$ ) present in different phonocardiographic recordings, concatenated for graphical purposes. The complete cardiac cycle has an approximate period of 800 ms, where the average duration of normal heart sounds is 70–150 ms for S1 and 60–120 ms for S2.

Since the range of frequencies that heart sounds generate are near the lowest threshold of sensitivity for the human ear, a practising physician requires extensive training guided by experienced physicians to make a correct diagnosis. In an alternative scenario, heart sounds can be recorded using electronic stethoscopes, allowing the practitioner to listen to heart sounds as many times as necessary to train the ear. This has proven effective in improving physicians' skills without the need to rely on available patients during hospital rotation [8]. In both cases, the diagnosis will depend fundamentally on the skill of the physician, subject to human error.

To alleviate the need for the correct training of the practising physician and to generate tools that can be used as auxiliary techniques in the diagnosis of heart valve diseases, in the present work, we implement a deep learning algorithm for the classification of common valvular diseases based on their temporal and spectral features. The dataset used contains phonocardiographic (PCG) records previously labelled in five classes: normal (N), aortic stenosis (AS), mitral regurgitation (MR), mitral stenosis (MS) and mitral valve prolapse (MVP).

It is worth mentioning that the previously described dataset has already been used to address similar classification problems. Firstly, the authors of the database, in 2018 [9], proposed to perform a multiclass classification comparing the performance of Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and a Deep Neural Network (DNN), and the extracted features were Discrete Wavelet Transform (DWT) and Mel Frequency Cepstral Coefficients (MFCC).

In the following two years, five more classification works were published using the same database, proposing a variety of feature extraction techniques and classification algorithms: (i) a set of eleven statistical features from the estimated instantaneous frequency of non-segmented PCG signals, for the binary classification and multiclassification comparing both random forest and KNN classifiers [10]; (ii) time-frequency features from the segmented cycles of the PCG signal through the wavelet synchrosqueezing transform, for a multiclass classifier for four of the five classes (MR, MS, AS, N) using a random forest classifier [11]; (iii) centroid frequency as the main feature to compare the performance of KNN and SVM, on both binary and multiclass problems [12]; (iv) local energy and local entropy features of the Chirplet transform from the PCG signal for a multiclass classifier for four of the five classes (MR, MS, AS, N) using a composite classifier [13]; (v) raw data and a WaveNet neural network for a multiclass classification model was proposed, to classify the PCG into the five classes [14]. All the above-mentioned works are briefly summarized in Table 3.

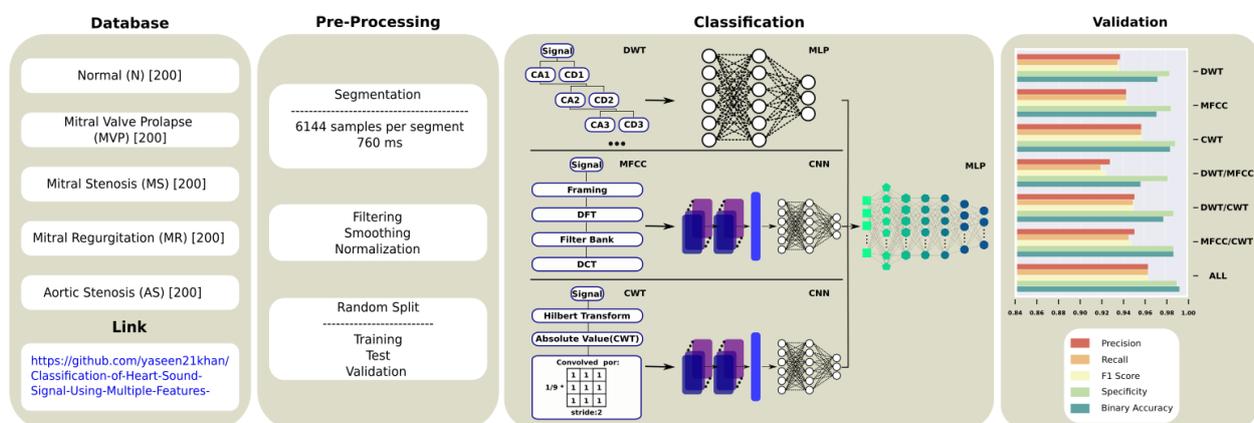
The present work stands out by using a deep learning algorithm that exploits frequency dynamics present throughout the cardiac cycle to differentiate between normal and

abnormal cardiac conditions. In particular, there are three properties that place this work as a new methodological proposal:

- (i) The use of Convolutional Neural Networks (CNN), together with the Continuous Wavelet Transform (CWT) and MFCCs, turning the time series classification problem into image classification, which has not been reported before for HVD classification, since all the related works use a vector-like feature arrangement or the raw time series.
- (ii) A transfer learning approach, which allowed us to migrate the deep learning algorithm previously trained for multiclass classification into a binary classification, permitting us to test the pre-trained model in a new classification problem, ensuring that the learning process generalizes and does not memorize, seeming quite flexible when compared.
- (iii) The complete model of the deep learning algorithm consists of two main stages. The first is formed by three parallel neural networks, each processing one of the three different features for individual cardiac cycles. The second is constituted by a perceptron that integrates the outputs of the previous stage, finding new patterns between them, improving the efficiency and effectiveness compared with the related works.

## 2. Materials and Methods

The classification of HVDs through the analysis of PCG signals consists of several stages, as shown in Figure 2: (i) generation of the dataset and labelling; (ii) pre-processing, which consists of cleaning the dataset, filtering the signal and segmentation by time windows; (iii) feature extraction; (iv) classification through a deep learning algorithm; (v) validation of the classifier when comparing predictions against the true labels.



**Figure 2.** Block diagram of the complete process.

The present work addresses the problem from the pre-processing. The signals were obtained from an open dataset [9] that consists of 200 records per class. These were digitized with a sampling frequency of 8 kHz, where each record has a duration of at least one second. To maintain uniformity in the data during the analysis, each record was segmented (when possible) into two windows of 6144 data points (0.768 s), which contain at least one complete cardiac cycle  $X_n$ . The details of the dataset after completing the segmentation process are shown in Table 1.

**Table 1.** Details of the dataset.

Binary Labels	Multiclassification Labels	No. of Recordings
Normal	Normal (N)	400
Abnormal	Aortic Stenosis (AS)	400
	Mitral Regurgitation (MR)	400
	Mitral Stenosis (MS)	380
	Mitral Valve Prolapse (MVP)	400
Total		1980

As can be seen from the results, the imbalance in the MS class does not affect the multiclass classifier performance. Given the optimization problem of any supervised model, where the size of the training data is related to the number of parameters of the model, we expect satisfactory training and good convergence [15–19].

### 2.1. Feature Extraction Techniques

Within the area of computer science, artificial intelligence and machine learning, feature extraction consists of the implementation of techniques aimed at extracting informative and non-redundant parameters from the measured data. Thus, the learning and generalization stage is facilitated.

Given the nature of the information in the records that shapes the dataset used in the present work, we decided to use three different techniques to extract the spectral features of the signals: MFCCs, CWT and energy from the DWT. The pseudocode for each of the feature extraction techniques is given in Appendix A.

#### 2.1.1. Mel Frequency Cepstral Coefficients

This is a state-of-the-art algorithm based on the known sensitivity variation of the critical frequency bandwidths present in the cochlea of the human ear, most used for the extraction of spectral information from time series [20]. Its success is due to its ability to represent the width of the spectrum of a signal in a compact form. Here, we will review each step in the process of creating the MFCCs, but motivated by considerations that adjust to the heart sound rates, where the summary of the stages is shown in Figure 6 inside the pink box [21].

**Framing:** The general purpose of this stage is to prepare the signal, segmenting it into  $N$  samples, where the segmentation window has a shift of  $m$  and the separation of the adjacent frames has no overlap. Subsequently, to analyze the frequency changes, the Discrete Fourier Transform is applied to each segment. Usually, 20–40 ms windows with a 50% overlap ( $\pm 10\%$ ) are used to process speech. However, frequencies lower than those present in the voice predominate in heart sounds, so it is proposed to use 64 ms windows without overlap.

**Discrete Fourier Transform (DFT):** The energy contribution of each of its frequency constituents is calculated for each segment through the DFT. This transformation of domains from time to frequency is described by [22]:

$$\mathcal{F}(x_n) = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi jkn}{N}}, k = 0, \dots, N-1 \quad (1)$$

whose resolution in frequency is given by  $fs/N$ , where  $fs$  is the sampling frequency,  $N$  the number of samples and  $|\mathcal{F}|$  the magnitude.

**Filter Bank:** Although the number of filters is a hyperparameter, the central frequency of each of them is calculated linearly from the theoretical maximum frequency recorded in

the signal. The first step consists of changing the scale from Hertz to Mels for the theoretical maximum frequency of the signal, according to the Nyquist Theorem [23]:

$$Mel(f) = 2595 \times \log_{10} \left( 1 + \frac{f}{700} \right) \tag{2}$$

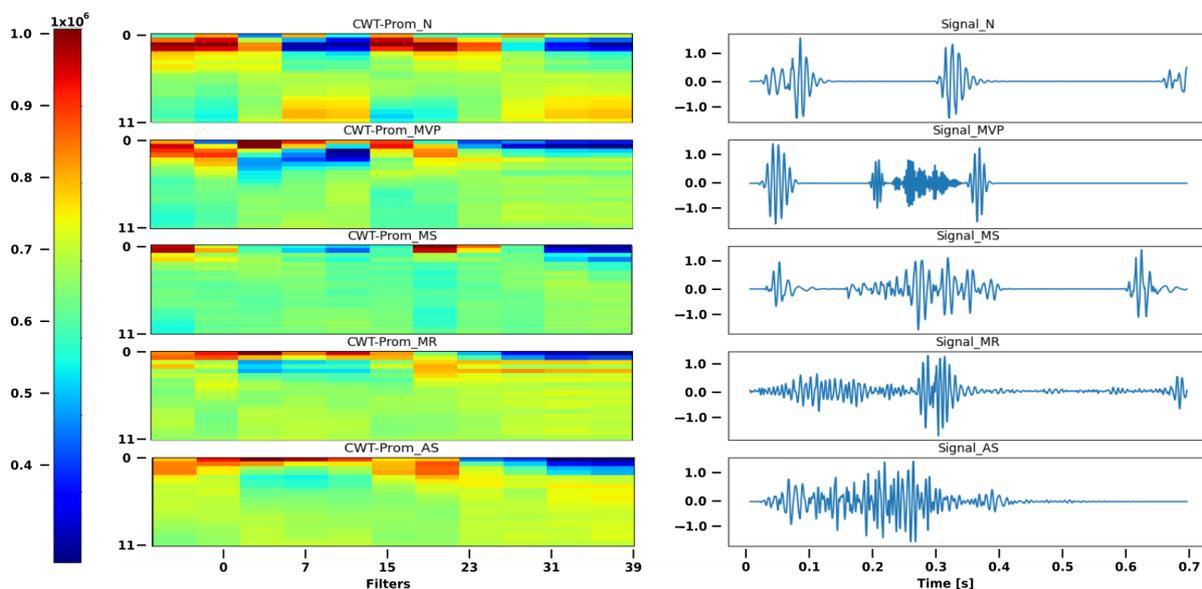
Once we have the theoretical maximum frequency in Mels,  $M$  equidistantly spaced frequencies are proposed to be used as central frequencies for the filters. Subsequently, it is necessary to remap the central frequencies from Mels to Hertz to design the  $H_m[k]$  filters [24] :

$$H_m[k] = \begin{cases} 0 & \text{if } k < f[m - 1] \\ \frac{2(k - f[m - 1])}{(f[m + 1] - f[m - 1])(f[m] - f[m - 1])} & \text{if } f[m - 1] \leq k < f[m] \\ \frac{2(f[m + 1] - k)}{(f[m + 1] - f[m - 1])(f[m] - f[m - 1])} & \text{if } f[m] \leq k \leq f[m + 1] \\ 0 & \text{if } k > f[m + 1] \end{cases} \tag{3}$$

Then,  $\mathcal{F} \times H_m$  is computed, obtaining a reduced spectral representation of the signal. Discrete Cosine Transform: Finally, it is necessary to apply a discrete cosine transform in order to obtain the MFCCs of each window, described as [25]:

$$F_k = 2 \frac{C_k}{N} \sum_{j=0}^{N-1} f(j) \cos \left[ \frac{2(j + 1)k\pi}{2N} \right], \begin{cases} c_k = \frac{1}{\sqrt{2}} & \text{if } k = 0 \\ c_k = 1 & \text{if } k > 0 \end{cases} \tag{4}$$

As mentioned, this process of multiplying the coefficients obtained after applying DFT by the filter bank is repeated for each segment of the signal, obtaining, as shown in Figure 3, an array with dimensions  $N \times M$ , where  $N$  is the number of windows and  $M$  the number of filters.



**Figure 3.** Results after applying MFCCs: The graph on the left shows, in a colorimetric scale, the value of the MFCCs averaged for each of the class sets. The graph on the right shows a normalized PGC record, randomly selected from each of the classes.

### 2.1.2. Continuous Wavelet Transform

Similar to the Short-Time Fourier Transform (STFT), the wavelet transform (WT) in both its continuous (CWT) and discrete (DWT) forms is a spectral decomposition. The easiest way to understand the WT is by comparing it with the STFT: for its part, the FT consists of decomposing the signal  $f(t)$  into its harmonic components in the form of sines and

cosines, while the WT decomposes the signal in the form of mother wavelets with different amplitudes and displacements. The mother wavelets are an effectively limited waveform in duration, with an average equal to zero. The mother wavelet used in the CWT was a Morlet, described by

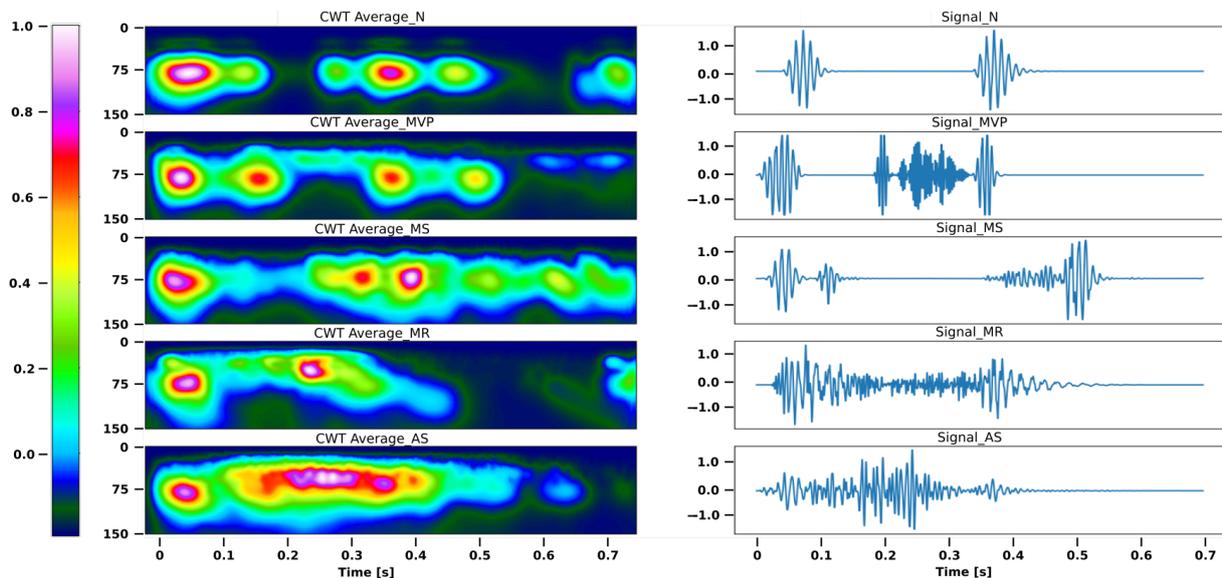
$$\psi(t) = e^{-\pi t^2} e^{i\pi t} \tag{5}$$

The CWT of a signal  $f(t)$  is given by [26]:

$$CWT(a, b) = \langle f, \psi_{a,b} \rangle \sum_0^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \tag{6}$$

Therefore, the integral is solved for  $a, b$  (scaling and shifting parameters), which performs a transformation of the signal  $f(t)$  from the time domain to a function in the time domain and scale.

By applying the CWT, we obtain a matrix representation of the coefficients of size  $N \times M$ , as shown in Figure 4, where  $N$  is the number of temporary points, while  $M$  is the number of scales used. Given the high dimensionality of these matrices, it was necessary to apply an averaging  $3 \times 3$  filter to reduce the computational demand.



**Figure 4.** CWT results: The graph on the left shows, in a colorimetric scale, the value of the coefficients obtained for each scale at each time point, averaged for each of the class sets, with 150 scales. The graph on the right shows a normalized PCG record, randomly selected from each of the classes.

**Hilbert Transform:** It is known that the multiresolution framework of the CWT can be improved through the Hilbert transform [27] if this is applied to the signal *a priori*. The fundamental reason that this transformation can be integrated with the wavelet transform is due to its scale and translation invariance, and its energy-conserving (unitary) nature [28].

Formally, the Hilbert transform  $\hat{s}(t)$  of a function  $s(t)$  is obtained by calculating the convolution of  $(s(t) \otimes 1/(\pi t))$  defined as [29]:

$$\hat{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau \tag{7}$$

This transformation has been shown to improve the performance of the neural network used, so it was implemented as part of the algorithm, as shown in the yellow box in Figure 6.

### 2.1.3. Discrete Wavelet Transform

As its name implies, the DWT is any WT where its wavelet function is discretized and, as with the CWT, captures the information in a time domain and scale.

DWT uses a family of orthonormal wavelets with unit energy, described by

$$\psi_{a,b}(t) = 2^{-a/2}\psi(2^{-j}t - b) \tag{8}$$

where, as in the CWT,  $a$  is the scaling parameter and  $b$  the translation parameter of the mother wavelet  $\psi$ , being Daubechies 4 the mother wavelet used in this algorithm. To analyze the spectral resolution of the data, it is necessary to enter the scaling function defined as [30]:

$$\phi(t) = \sum_{k=-1}^{N-2} (-1)^k C_{k+1}\psi(2t + k) \tag{9}$$

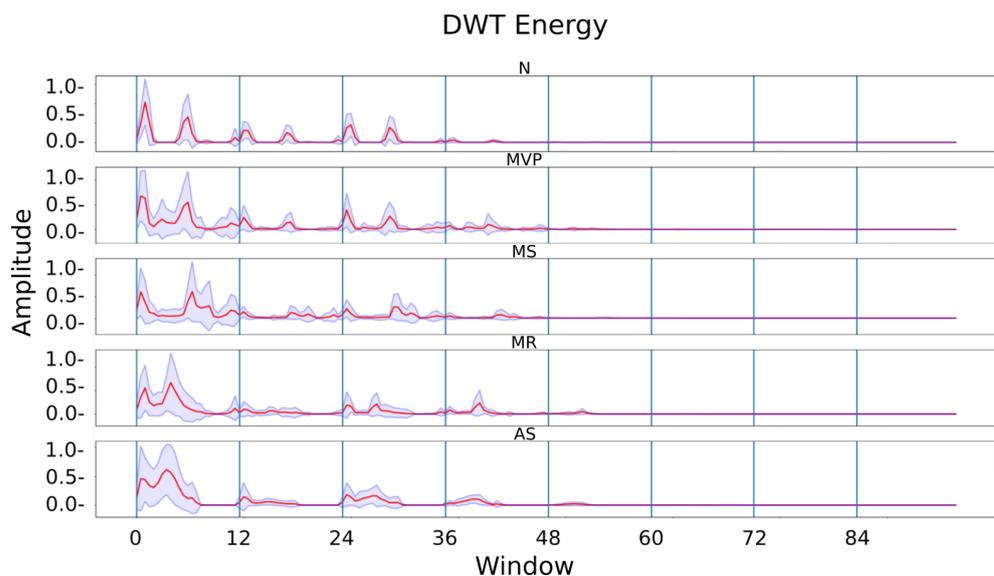
The  $C_k$  are known as wavelet coefficients. These coefficients are arranged as a transformation matrix that is applied to a vector of data. In this way, the coefficients are stated in two different patterns, one that works as a fading filter (low-pass filter) and the other as a pattern that shows only the details of the information (high-pass filter).

This concept of analyzing a signal using filter banks is known as Mallat tree decomposition. Inside the green box in Figure 6, we can see how a signal  $x(n)$  is decomposed into approximations  $a_j(n)$  and details  $d_j(n)$  by the effect of the filters  $h_j(n)$  and  $g_j(n)$ .

Once the multiresolution analysis has been applied, the energy of each detail and the last approximation  $x[n]$  is calculated, defined as:

$$E_x = \sum_{-\infty}^{\infty} |x[n]|^2 \tag{10}$$

As in MFCCs, the signal was segmented into 12 windows of 512 data points, where the energy was calculated by the level of decomposition. In Figure 5, the average energy and its standard deviation of each segment per class are shown.



**Figure 5.** Energy calculation: The red line and the blue shadows are the averages and their standard deviation for each record per class, respectively. The vertical lines separate the decomposition levels in the following order: Approximation 7 (A7), Detail 7 (D7), D6, D5, D4, D3, D2, D1.

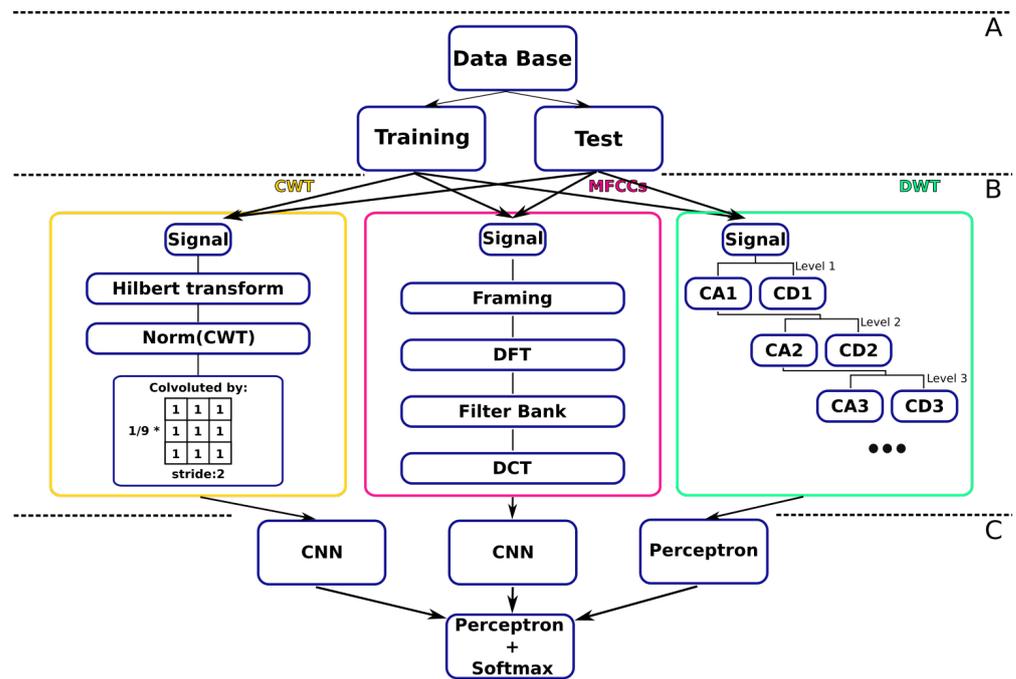
### 2.2. Classification Algorithms

A classifier is an algorithm that assigns a putative class to a pattern, this being the identifier of an object, signal or phenomenon encoded based on its features or attributes [31], which in turn may be continuous, categorical or binary. If instances are given with known labels (the corresponding correct outputs), then the learning is called supervised, in contrast to unsupervised learning, where instances are unlabelled. By applying these unsupervised (clustering) algorithms, researchers hope to discover unknown, but useful, classes of items [32].

For the classification and recognition of normal and abnormal PCG signals described above, a supervised deep learning algorithm was implemented, which refers to a multilayer directed model with a non-linear topology and multiple inputs and/or outputs, capable of extracting by itself the patterns that define each class.

### 2.3. Deep Learning (DL)

The DL used in the present work consists of two main stages, as shown in Figure 6. The first stage consists of three parallel artificial neural networks, aimed at generalizing patterns: a multilayer perceptron network whose input was the coefficients obtained from the energy calculation after spectral decomposition applying DWT, and two convolutional neural networks (CNN) that, as input, receive, respectively, matrices with the MFCCs and CWT coefficients. In the second stage, the outputs of the three independent networks were concatenated to feed a multilayer perceptron that gives us the probability that the PCG signal belongs to any of the classes or subclasses mentioned in Table 1.



**Figure 6.** DNN block diagram: The diagram shows the proposed architecture. (A) Partition of the dataset into the training and test sets is exemplified. (B) Feature extraction techniques. (C) Classification algorithms.

The main difference in the multiclass or binary classification consists of the number of neurons in the last layer, being two or five for the binary or multiclass classification, respectively. However, only the multiclass network was trained, and to perform the binary classification, a transfer learning approach was implemented. This implies that, once the multiclass model was trained, only the last layer was changed. The labels of the abnormal class for the binary classification are split into four subclasses for the multiclass classification,

and the activation function for the output layer differs according to the problem. For the multiclass classification, a probabilistic activation function is used, described as

$$O_i = \frac{e^{y_i^l}}{\sum_{k=1}^K e^{y_k^l}} \quad (11)$$

while, for the binary classification, the activation function used was a sigmoid, described as

$$O = \frac{1}{1 + e^{-x}} \quad (12)$$

The feature extraction algorithms, the DL algorithm and its constituent networks were implemented on Python 3.9. In particular, the DL neural network was designed using Keras 2.4.3 on Ubuntu 20.04 distribution. Keras is a high-level neural network library written in Python, which can run on top of TensorFlow or Theano. The model and the respective sub-architectures were allowed to run on an Intel processor (i7-9700) with 32 GB of RAM and on an NVIDIA GeForce GTX 1650 GPU with 4 GB of VRAM memory. The pseudocode describing the architecture for the DL algorithm proposed is presented in Appendix B.

### 2.3.1. Convolutional Neural Network (CNN)

A Convolutional Neural Network is a bio-inspired algorithm that attempts, to some extent, to emulate the ability of the visual cortex to extract key features from images and process them. A CNN has multiple layers, including a convolutional layer, normalization, non-linearity layer, max-pooling and a multilayer perceptron.

The discrete convolution between the filter and the coefficient matrix is mathematically defined as:

$$\text{conv}(I, K)_{x,y} = \sum_{i=0}^{n_{f1}-1} \sum_{j=0}^{n_{f2}-1} K_{(i,j)} I_{(x+i,y+j)} \quad (13)$$

It is possible to deduce that, if the image dimension is given by  $(n_H, n_W)$  and the filter dimension is given by  $(f_1, f_2)$ , the dimension of the convolution will be

$$\text{dim}(\text{conv}(I, K)) = \left[ \frac{n_H - f_1}{s} + 1, \frac{n_W - f_2}{s} + 1 \right] \quad (14)$$

Max-pooling is a particular case of a convolutional layer where the filter is a matrix of ones and, after the convolution, a maximum function is applied. By convention, we consider a square filter with dimensions  $f_1 = f_2 = 2$  and  $s = 2$ . This operation is defined as:

$$\max(K_{(i,j)} I_{(x+i,y+j)}) \quad (15)$$

In CNN terminology, the pair of convolution and max-pooling layers in succession is often referred to as a convolutional layer [33].

### 2.3.2. Multilayer Perceptron

A Perceptron [34] refers to a fully connected artificial neural network, with more than one hidden layer, that takes as input the output of the previous layer, multiplies it by a vector of synaptic weights, adds the result and passes it through an activation function:

$$y_i^l = \sigma \left( \sum_j y_j^{(l-1)} w_{(j,i)}^l + w_{(0,i)}^l \right) \quad (16)$$

In this matrix notation,  $y_i^l$  denotes the output of the  $i$ -th neuron of the  $l$ -th layer,  $w_{j,i}^l$  are the synaptic weights of the neurons and  $\sigma(x)$  is the activation function.

The optimization algorithm used to update the synaptic weights  $W$  (training) was the descending gradient:

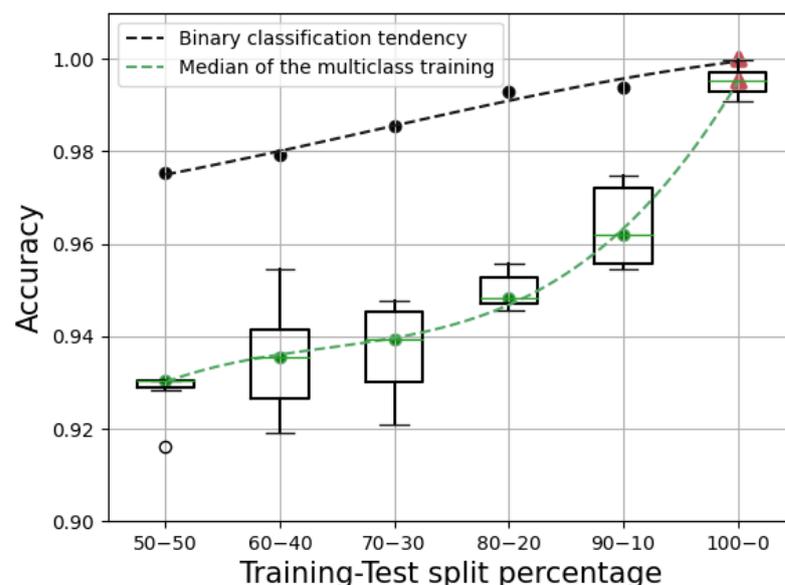
$$W \leftarrow W - \alpha \frac{\partial MSE}{\partial W} \quad (17)$$

where  $W$  is the synaptic weight matrix,  $\alpha$  is the learning rate, and  $MSE$  is the mean square error.

### 3. Results

During the classifier construction, the experimentation consisted of designing independent classifiers that, as input, had the features extracted from the dataset. The DL architecture for the classification and prediction of the normal and abnormal heart sounds has three neural networks that, independently and in parallel, generalize the patterns that define the classes from the features mentioned in Section 2. Each constitutive network was tested independently and in pairs to find the most competitive configuration.

As mentioned before, the models were initially trained to perform a multiclass classification, and later, a transfer learning approach was applied to modify the last layer of the trained network to obtain a binary classification. In Figure 7, the boxplot shows the F1 score obtained by testing the complete multiclass model 10 times, while the black dotted line illustrates the tendency of the binary accuracy obtained of the best model, for percentages of the dataset randomly split into training and testing sets, ranging from 50–50% to 100–0%.

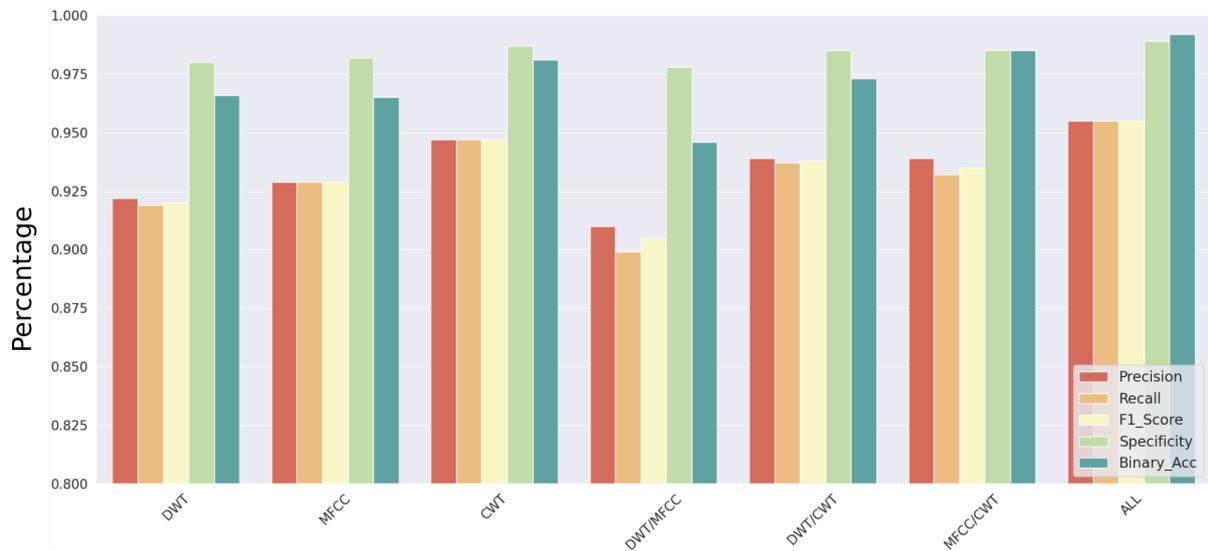


**Figure 7.** Performance of the complete model for different percentages of training data split. Boxplot of F1 score of a ten folds test; green dotted line shows the increase in the median adjusted to a quadratic function. The black dotted line shows the tendency of the best model for each percentage transferred to classify binary classes, and the red triangles account for the performance reached with 100% of the dataset as the training set.

Ignoring the (almost) perfect performance of the model achieved using the entirety of the dataset as the training set, it is possible to observe that, although the training data percentage increased, the performance for the multiclass classification has an increase adjusted to a quadratic function. The performance of the binary classification reaches a plateau when training with 80% of the data. Therefore, the complete model and submodels were trained and tested using an 80–20% split of the dataset, respectively.

Figure 8 summarizes the performance of the independent models, pairwise combinations of them and the complete model, previously trained and evaluated using the testing data, measured through precision, recall, F1 score and specificity for multiclass classifi-

ation, and accuracy for the binary classification. To prevent overtraining, intermediate dropout [35] layers were implemented in each of the networks, with a uniform probability of deactivation of 30% for each of the neurons.



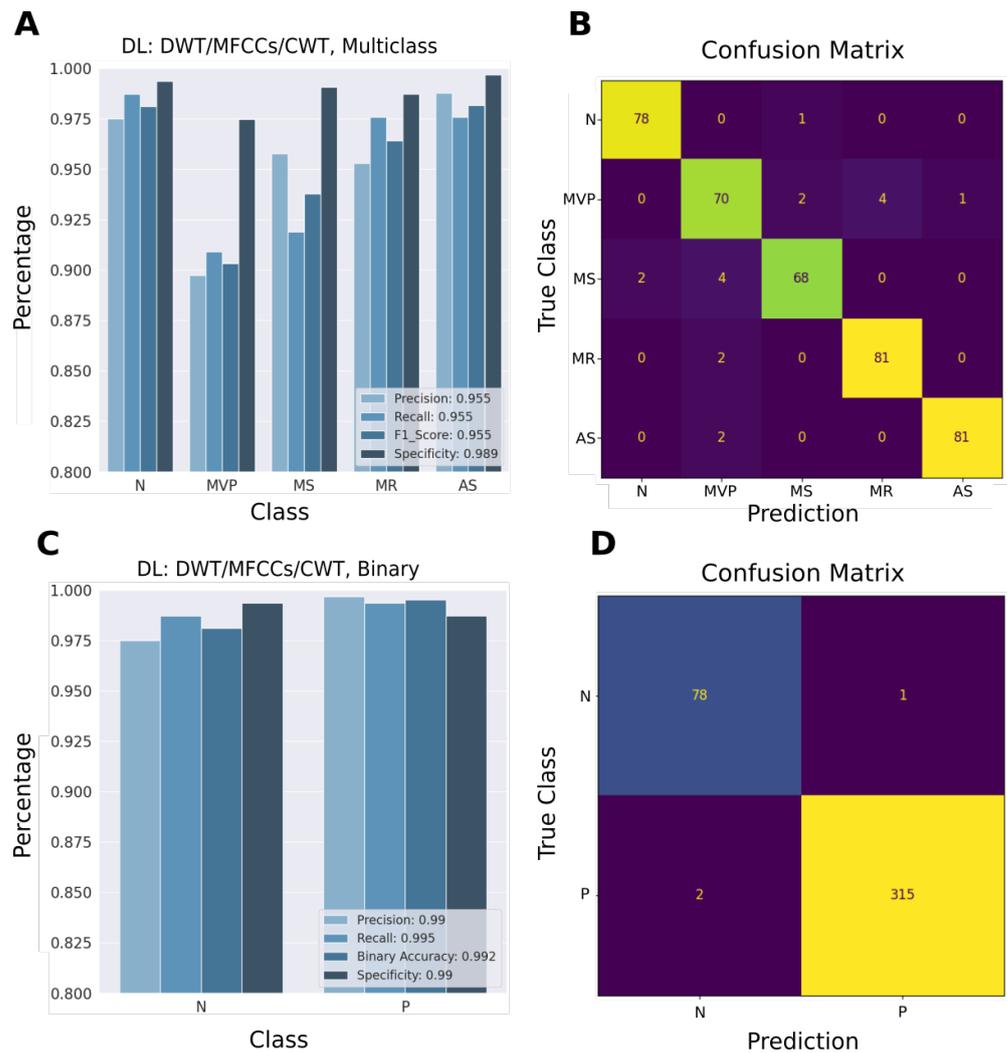
**Figure 8.** Models’ performance: The graph shows the precision, recall, F1 score, specificity and binary accuracy achieved by each proposed model: the independent networks, a combination of pairs of them and the complete model based on Table 2.

**Table 2.** Comparative table between precision, recall, F1 score and specificity achieved by each proposed model: the independent networks, a combination of pairs of them and the complete model.

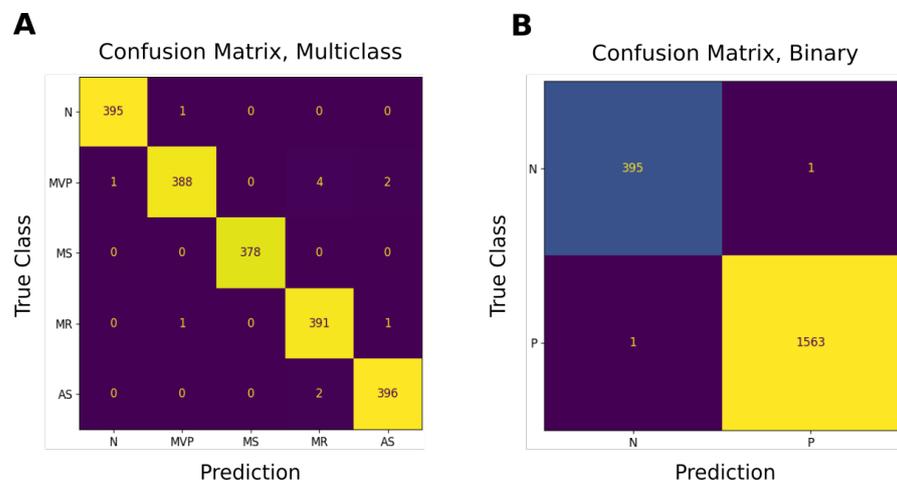
	Precision	Recall	F1 Score	Specificity	Binary Accuracy
DWT	0.922	0.919	0.92	0.98	0.966
MFCC	0.929	0.929	0.929	0.982	0.965
CWT	0.947	0.947	0.947	0.987	0.981
DWT/MFCC	0.91	0.899	0.905	0.978	0.946
DWT/CWT	0.939	0.937	0.938	0.985	0.973
MFCC/CWT	0.939	0.932	0.935	0.985	0.985
ALL	0.955	0.955	0.955	0.989	0.992

After training and testing the networks, it is possible to observe that the configurations with the best performance are those that have as inputs both of the spectral features extracted by MFCCs and CWT. On the other hand, if we compare the performance of the networks that only have one of the features as input, we can observe that the CNN trained with the features extracted by CWT has the best performance.

Figure 9A,C show the detailed F1 score, precision, recall and specificity obtained for each class in the complete architecture using 20% of the dataset as the testing set. It is possible to observe that 95% of the classes were correctly classified in a multiclass classification, according to their F1 score, and 99% in a binary classification according to the binary accuracy. Furthermore, the confusion matrix, on which all the metric calculations were based, is shown in Figure 9B,D, where each column of the matrix represents the number of predictions of each class, while each row represents the instances in the true class. Meanwhile, Figure 10A,B show the confusion matrices obtained for each class in the complete architecture, using 100% of the dataset as the training/testing set in a multiclass and binary classification, respectively.

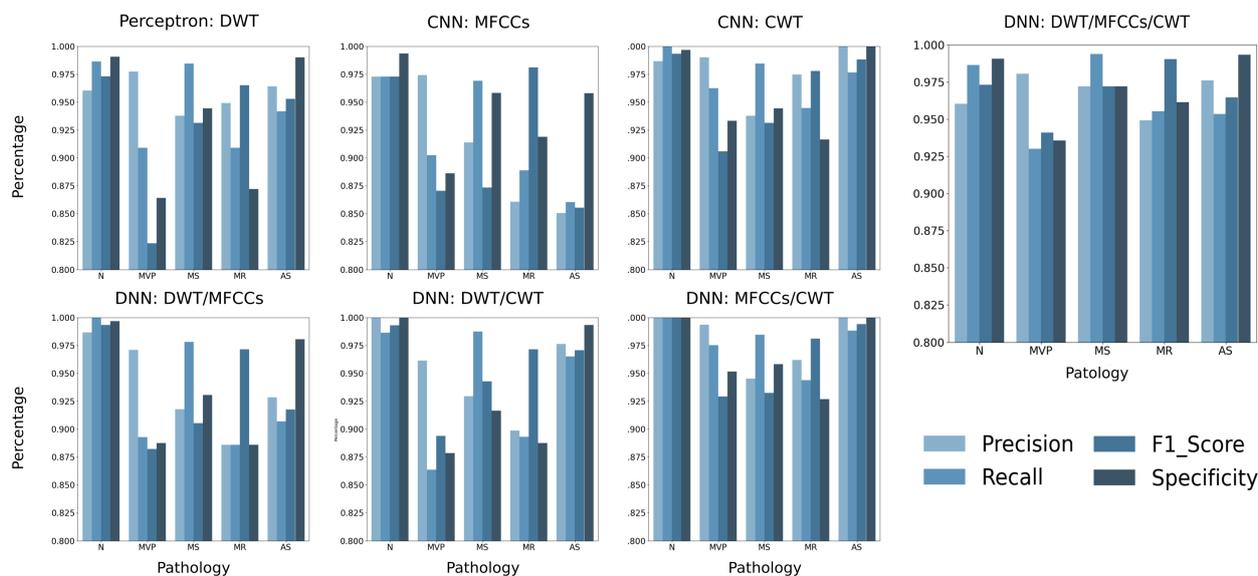


**Figure 9.** DL performance results (20% of dataset as test set): (A,C) Precision, recall, F1 score and specificity of the complete model for the testing set in a multiclass and binary classification, respectively. The vertical axis shows only the percentage from 0.8 to 1.0 to facilitate the visualization of the results. (B,D) Confusion matrices for multiclass and binary classification, respectively.



**Figure 10.** DL performance results (20% of dataset as training/testing set): (A,B) Confusion matrices for multiclass and binary classification, respectively.

Finally, Figure 11 shows the performance of each network and the combinations used for each of the classes evaluated through the same metrics. It is possible to observe that although the performance varies between the networks, the results are consistent between each of the proposed configurations, where the precision and recall for each class are spliced. Along with the intermediate dropout layers, these results confirm that the networks were correctly generalizing the features that define each class and not memorizing the patterns for each signal.



**Figure 11.** Experimental results: The bar graphs show the results achieved by each architecture, for each class, measured using precision, recall, F1 score and specificity as metrics.

#### 4. Discussion

Since PCG signals contain information about the condition of the heart valves, the analysis and development of technology that allows the correct classification of these signals can be used as an auxiliary technique for the precise and timely diagnosis of HVDs. The information of the PCG signals has a temporality, relative intensity and spectral encoding, which invites us to use feature extraction techniques to enhance their contributions, facilitating their classification. To propose a novel method to detect HVDs, we decided to build a DL neural network that would receive as input a set of features extracted by CWT, MFCCs and the energy associated with each level of decomposition by DWT.

Through the experimentation, the performance of each constituent part of the network was measured independently and in pairs for a multiclass and binary classification problem. The configurations with the best performance were those that generalized from the features extracted by MFCCs and CWT, scilicet, the complete or partial DNN model that consisted of two parallel CNNs. Nevertheless, the F1 scores and binary accuracy of the complete model reached values slightly higher than 95% and 99%, respectively, indicating that the energy associated with each level of decomposition by DWT contains information that CWT or MFCCs does not, making the complete model the most competitive architecture.

In addition to the F1 score, one of the principal metrics to evaluate the competitiveness of the proposed model is the specificity, since this is the only one that takes into account the false positives of each class. In particular, it is of great interest to observe the specificity calculated in the “Normal” class, since a perfect score (100%) in this metric would ensure that no patient with any HVD would be classified as a healthy person, which is the worst-case scenario during diagnosis, and could mean a risk for the patient’s health. We consider that, for clinical applications, this is the most important metric that must be accounted for and where this work stands out, having specificities in the “Normal” class of up to 99%.

It should be noted that this work is not the first to address a similar problem; however, a fair comparison can only be made with those publications that have used the same database as us. Table 3 shows, in a comparative way, seven of these works (including the present one), which were given the task of applying artificial intelligence techniques to perform binary or multiclass classification.

If we take the performance of the model proposed in the present work, achieved using 80% of the dataset as the training set and measured through the F1 score, and compare it with the F1 score of the work proposed by [9], or any global accuracy proposed in the other works, it would seem that our proposal is less competitive. However, there are certain neglected considerations in the mentioned works, which the present work addresses more thoroughly.

First, it is possible to observe that the results reported in [9,12,14] are those obtained during training, since the sum of the data contained in the confusion matrices is the total of the data in the dataset. This has important implications when interpreting their results, since using 100% of the dataset during training allows us only to know the degree of compaction of the data and the memorization capacity of the classification algorithms, but it is not possible to measure the capability of the algorithm to generalize, as it has not been tested with unseen data.

In order to compare our work with the state-of-the-art algorithms mentioned in Table 3, the performance of our model using 100% of the dataset as the training set was experimentally calculated, surpassing all, in both multiclass and binary classification. Nevertheless, we do not encourage the use of this approach since it is not a correct computational practice [31].

Another neglected point is the imbalance of the classes present during the binary classification reported [10,12]. In the first one, it is reported that the data are separated into training (70%) and testing (30%) sets. In both cases, the number of data present in the “abnormal” class is several times greater than that present in the “normal” class. This tends to skew the classifier by having more information about a class, increasing the probability of an incorrect classification.

It seems that the proposals with which we can appropriately compare our work are [11,13]. Nonetheless, they arbitrarily remove the MVP class from the database, and we can only make assumptions about why: as mentioned by Delling et al. in 2016 [36], MVP progresses into MR over a period of 3 to 16 years in one fourth of individuals. Since MVP is the precursor of MR, it is safe to suppose that both share similarities, so much so that, in our results, the MVP is the worst classified pathology, mostly confused with MR.

Finally, most of the works chose to measure the performance of their classifiers through global accuracy and not the F1 score. This is only a problem when classes are highly imbalanced, as the accuracy does not take into account how the data are distributed. Since, in most real-life classification problems, an imbalanced class distribution exists, the F1 score is a better metric to evaluate our model on. Nevertheless, to make a fair comparison, we also calculate the global accuracy in our model, which surpasses all previously mentioned works, even under the 80% training set condition.

Given all the aforementioned considerations, and the scrutiny under which this work was developed, it is possible to conclude that the results achieved by our algorithm are highly competitive and reliable compared to those present in the state of the art that use the same database (Table 3). Nevertheless, it is necessary to test new, non-spectral feature extraction techniques, increasing the attributes for the improvement of the proposed model.

**Table 3.** Comparative table between works that used the same dataset. The DL algorithm that is proposed in the present work is mentioned twice, showing the results using 80% or 100% of the dataset as the training set. When more than one classifier was used, the best one (\*) is presented.

Author	Feature Extraction	Classifier	Classification	Precision	Recall	F1 Score	Specificity	Binary Accuracy	Global Accuracy
Son et al., 2018 [9]	DWT and MFCCs	SVM *, KNN, DNN	Multiclass	–	98.2%	99.7%	99.4%	–	97.9%
Alqudah, A. M., 2019 [10]	Eight statistical moments from the Instantaneous Frequency Estimation + PCA	KNN * and Random Forest	Binary	100%	98.28%	–	100%	99.60%	94.8%
			Multiclass	95.12%	94.78%	–	98.7%		
Ghosh, S. K. et al., 2019 [11]	13 statistical moments from the Wavelet Synchrosqueezing Transform	Random Forest	Multiclass	–	–	–	–	–	95.13
Upretee, P., and Yuksel, M. E., 2019 [12]	Centroid Frequency Estimation	SVM and KNN *	Binary	99.6%	99.76%	–	98.83%	99.75%	96.5%
			Multiclass	96.50%	96.5%	–	99.12%		
Ghosh, S. K., et al., 2020 [13]	Local energy and entropy from the Chirplet transform	Multiclass composite classifier	Multiclass	98.0%	98.1%	–	99.3%	–	98.33%
Oh, S. L. et al., 2020 [14]	Raw Data	WaveNet	Multiclass	94.0%	–	–	99.25%	–	97.0%
			Binary	99%	99.5%	–	99%	99.2%	99.2%
Present Work (80% Training)	DWT, MFCCs, CWT	DL neural network	Multiclass	95.5%	95.5%	95.5%	98.9%	–	98.5%
			Binary	99.8%	99.8%	–	99.8%	99.8%	99.8%
Present Work (100% Training)	DWT, MFCCs, CWT	DL neural network	Multiclass	99.3%	99.3%	99.3%	99.8%	–	99.7%

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

**Funding:** This research was funded by the Instituto Politécnico Nacional through the projects SIP20201657 and SIP20210473.

**Data Availability Statement:** The data supporting the reported results can be found at <https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features-> (accessed on 27 January 2022).

**Acknowledgments:** We thank CONACyT for partial support of the present work.

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

## Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DNN	Deep Neural Network
DWT	Discrete Wavelet Transform
HPF	High-Pass Filter
HVD	Heart Valve Disease
LPF	Low-Pass Filter
WHO	World Health Organization

## Appendix A. Feature Extraction Techniques Algorithmic Expression

---

### Algorithm A1 MFCC Algorithm

---

**Input:** Cardiac cycle  $X_n$ , and the number of filters  $k$ .

**Output:**  $M$  matrix, containing the frequency dynamics of  $X_n$ , every  $m$  bins, without overlap.

- 1: Segment  $X_n$  into  $N$  windows of constant length  $m$ , with no overlap.
  - 2: Define the maximum theoretical frequency for the windows in Hz  $f$ .
  - 3: Calculate  $k$  equidistant central frequencies from zero to  $f$  in Mels.
  - 4: Generate the filter bank  $H_m[k]$  (Equation (3))
  - 5: **for** each  $i$  in  $N$  **do**
  - 6:      $\mathcal{F}$  = Apply DWT to  $i$  (Equation (1)).
  - 7:      $\mathcal{X}$  = Multiply  $\mathcal{F}$  by  $H_m[k]$ .
  - 8:     Calculate the power of  $\mathcal{X}$  in dB.
  - 9:      $F_k$  = Apply DCT to  $\mathcal{X}$  (Equation (4)).
  - 10:      $M_n = F_k$ .
  - 11: **end for**
- 

### Algorithm A2 CWT Algorithm

---

**Input:** Cardiac cycle  $X_n$ , scaling factors  $L_s, H_s$ .

**Output:**  $C$  matrix, containing the frequency dynamics of the cardiac cycle, each time bin.

- 1:  $\psi(t) \leftarrow$  Morlet mother wavelet (Equation (5)).
  - 2:  $\mathcal{H} = \text{Hilbert}(X_n)$  (Equation (7)).
  - 3: Define a number of subscales or *Voices*.
  - 4:  $Scales \leftarrow$  Array from the lowest to the highest scaling factor, such as  $([L_s:1:H_s])$ .
  - 5: **for**  $s$  in  $Scales$  **do**
  - 6:     **for**  $v$  in  $Voices$  **do**
  - 7:         Define subscales such as  $A = (2^{v+\frac{s}{v}})$
  - 8:     **end for**
  - 9: **end for**
  - 10: **for**  $a$  in  $A$  **do**
  - 11:      $\mathcal{X}(a) = \mathcal{H} \times \psi(a, t)$  (Equation (6)).
  - 12:      $C_a = |\mathcal{X}(a)|$
  - 13: **end for**
- 

### Algorithm A3 DWT Algorithm

---

**Input:** Cardiac cycle  $X_n$ , decomposition levels  $l$ .

**Output:**  $D$  array, containing the energy at each level of decomposition, in  $m$  bins, for  $X_n$ .

- 1: Segment  $X_n$  into  $N$  windows of constant length  $m$ , with no overlap.
  - 2: Define the low-/high-pass filter  $[h_j, g_j]$  coefficients (Equation (9)).
  - 3: **for**  $i$  in  $N$  **do**
  - 4:      $a_0 = \text{Downsample}(h_j(i))$
  - 5:      $d_0 = \text{Downsample}(g_j(i))$
  - 6:      $\text{Energy}[0] = \Sigma(|d_0|^2)$
  - 7:     **for**  $j$  in  $l - 1$  **do**
  - 8:          $a_{j+1} = \text{Downsample}(h_j(a_j))$
  - 9:          $d_{j+1} = \text{Downsample}(g_j(a_j))$
  - 10:          $\text{Energy}[j + 1] = \Sigma(|d_{j+1}|^2)$
  - 11:     **end for**
  - 12:      $\text{Energy}[l + 1] = \Sigma(|a_l|^2)$
  - 13: **end for**
  - 14:  $D = \text{Sort Energy}$  by level of decomposition.
-

## Appendix B. Classifier Algorithmic Expression

---

### Algorithm A4 DL Algorithm

---

**Input:** Dataset: Pre-segmented cardiac cycle of each patient  $X_n$ .

**Output:** Performance of the DL algorithm.

```

1: for each  $X$  in Dataset do
2:   MFCCs  $\leftarrow$  Append(call MFCCs( $X$ ))
3:   CWT  $\leftarrow$  Append(call CWT( $X$ ))
4:   DWT  $\leftarrow$  Append(call DWT( $X$ ))
5: end for

6: Split [MFCCs,CWT,DWT] into Training/Testing subsets.

7: CNN for MFCCs Architecture
8: input: MFCCs
9: CNN  $\leftarrow$  Convolution layer (32 Filters ( $3 \times 3$ ))
10: Batch Normalization + Nonlinear layer + MaxPooling
11: CNN  $\leftarrow$  Convolution layer (64 Filters ( $3 \times 3$ ))
12: CNN  $\leftarrow$  Flatten()(CNN)
13: MLP  $\leftarrow$  32 output neurons (CNN)
14: output1: pattern of 32 attributes.

15: CNN for CWT Architecture
16: input: CWT
17: CNN  $\leftarrow$  Convolution layer (16 Filters ( $21 \times 21$ ))
18: Batch Normalization + Nonlinear layer + MaxPooling
19: CNN  $\leftarrow$  Convolution layer (8 Filters ( $11 \times 11$ ))
20: Batch Normalization + Nonlinear layer + MaxPooling
21: CNN  $\leftarrow$  Convolution layer (4 Filters ( $7 \times 7$ ))
22: CNN  $\leftarrow$  Flatten()(CNN)
23: MLP  $\leftarrow$  32 output neurons (CNN)
24: output2: pattern of 32 attributes.

25: MLP for DWT Architecture
26: input: DWT
27: MLP  $\leftarrow$  Four layers(512, 128, 128, 64, neurons)
28: MLP  $\leftarrow$  32 output neurons
29: output3: pattern of 32 attributes.

30:  $P$   $\leftarrow$  concatenate outputs 1, 2 and 3 as a pattern array of 96 attributes
31: MLP for Classification
32: input:  $P$ 
33: MLP  $\leftarrow$  Five layers(96, 64, 32, 96, 16, neurons)
34: MLP  $\leftarrow$   $n$  output neurons (softmax or sigmoid) according to the number of classes
35: Train, test and evaluate DL algorithm

```

---

## References

1. WHO. WHO Reveals Leading Causes of Death and Disability Worldwide: 2000–2019—PAHO/WHO | Pan American Health Organization; WHO: Geneva, Switzerland, 2019.
2. Nkomo, V.T.; Gardin, J.M.; Skelton, T.N.; Gottdiener, J.S.; Scott, C.G.; Enriquez-Sarano, M. Burden of valvular heart diseases: A population-based study. *Lancet* **2006**, *368*, 1005–1011. [[CrossRef](#)]
3. Mondal, A.; Kumar, A.K.; Bhattacharya, P.; Saha, G. Boundary estimation of cardiac events S1 and S2 based on Hilbert transform and adaptive thresholding approach. In Proceedings of the 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT), Kharagpur, India, 28–30 March 2013; pp. 43–47.
4. Randhawa, S.K.; Singh, M. Classification of heart sound signals using multi-modal features. *Procedia Comput. Sci.* **2015**, *58*, 165–171. [[CrossRef](#)]
5. Chizner, M.A. Cardiac auscultation: Rediscovering the lost art. *Curr. Probl. Cardiol.* **2008**, *33*, 326–408. [[CrossRef](#)] [[PubMed](#)]

6. Liu, C.; Springer, D.; Li, Q.; Moody, B.; Juan, R.A.; Chorro, F.J.; Castells, F.; Roig, J.M.; Silva, I.; Johnson, A.E.; et al. An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* **2016**, *37*, 2181–2213. [[CrossRef](#)] [[PubMed](#)]
7. Varghees, V.N.; Ramachandran, K.I. A novel heart sound activity detection framework for automated heart sound analysis. *Biomed. Signal Process. Control* **2014**, *13*, 174–188. [[CrossRef](#)]
8. Tokuda, Y.; Matayoshi, T.; Nakama, Y.; Kurihara, M.; Suzuki, T.; Kitahara, Y.; Kitai, Y.; Nakamura, T.; Itokazu, D.; Miyazato, T. Cardiac auscultation skills among junior doctors: Effects of sound simulation lesson. *Int. J. Med. Educ.* **2020**, *11*, 107. [[CrossRef](#)]
9. Son, G.Y.; Kwon, S. Classification of heart sound signal using multiple features. *Appl. Sci.* **2018**, *8*, 2344.
10. Alqudah, A.M. Towards classifying non-segmented heart sound records using instantaneous frequency based features. *J. Med. Eng. Technol.* **2019**, *43*, 418–430. [[CrossRef](#)]
11. Ghosh, S.K.; Tripathy, R.K.; Ponnalagu, R.; Pachori, R.B. Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features. *IEEE Sens. Lett.* **2019**, *3*, 1–4. [[CrossRef](#)]
12. Upretree, P.; Yüksel, M.E. Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 24–26 April 2019; pp. 1–4.
13. Ghosh, S.K.; Ponnalagu, R.; Tripathy, R.; Acharya, U.R. Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals. *Comput. Biol. Med.* **2020**, *118*, 103632. [[CrossRef](#)]
14. Oh, S.L.; Jahmunah, V.; Ooi, C.P.; Tan, R.S.; Ciaccio, E.J.; Yamakawa, T.; Tanabe, M.; Kobayashi, M.; Acharya, U.R. Classification of heart sound signals using a novel deep WaveNet model. *Comput. Methods Programs Biomed.* **2020**, *196*, 105604. [[CrossRef](#)] [[PubMed](#)]
15. Chawla, N.V.; Japkowicz, N.; Kotcz, A. Special issue on learning from imbalanced data sets. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 1–6. [[CrossRef](#)]
16. Wang, B.; Japkowicz, N. Imbalanced Data Set Learning with Synthetic Samples. In Proceedings of the IRIS Machine Learning Workshop, Ottawa, ON, Canada, 9 June 2004.
17. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
18. Lin, M.; Tang, K.; Yao, X. Dynamic sampling approach to training neural networks for multiclass imbalance classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 647–660.
19. Wang, S.; Liu, W.; Wu, J.; Cao, L.; Meng, Q.; Kennedy, P.J. Training deep neural networks on imbalanced data sets. In Proceedings of the 2016 international joint conference on neural networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 4368–4374.
20. Gaikwad, S.K.; Gawali, B.W.; Yannawar, P. A Review on Speech Recognition Technique. *Int. J. Comput. Appl.* **2010**, *10*, 16–24. [[CrossRef](#)]
21. Rabiner, L.; Juang, B. *Fundamentals of Speech Recognition*; Pearson PLC: London, UK, 1993.
22. Heckbert, P. Fourier transforms and the fast Fourier transform (FFT) algorithm. *Comput. Graph.* **1995**, *2*, 15–463.
23. Umesh, S.; Cohen, L.; Nelson, D. Fitting the mel scale. In Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP99 (Cat. No. 99CH36258), Phoenix, AZ, USA, 15–19 March 1999; Volume 1, pp. 217–220.
24. Sigurdsson, S.; Petersen, K.B.; Lehn-Schiøler, T. Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music. In Proceedings of the ISMIR, Victoria, BC, Canada, 8–12 October 2006; pp. 286–289.
25. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete cosine transform. *IEEE Trans. Comput.* **1974**, *100*, 90–93. [[CrossRef](#)]
26. Sinha, S.; Routh, P.S.; Anno, P.D.; Castagna, J.P. Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics* **2005**, *70*, P19–P25. [[CrossRef](#)]
27. Chaux, C.; Duval, L.; Pesquet, J.C. Hilbert pairs of M-band orthonormal wavelet bases. In Proceedings of the 2004 12th European Signal Processing Conference, Vienna, Austria, 6–10 September 2004; pp. 1187–1190.
28. Chaudhury, K.N.; Unser, M. Construction of Hilbert transform pairs of wavelet bases and Gabor-like transforms. *IEEE Trans. Signal Process.* **2009**, *57*, 3411–3425. [[CrossRef](#)]
29. Johansson, M. The Hilbert Transform. Master’s Thesis, Växjö University, Växjö, Sweden, 1999; Volume 19. Available online: [http://w3.msi.vxu.se/exarb/mj\\_ex.pdf](http://w3.msi.vxu.se/exarb/mj_ex.pdf) (accessed on 1 May 2021).
30. Shensa, M.J. The discrete wavelet transform: Wedding the a trous and Mallat algorithms. *IEEE Trans. Signal Process.* **1992**, *40*, 2464–2482. [[CrossRef](#)]
31. Harrington, P. *Machine Learning in Action*; Simon and Schuster: New York, NY, USA, 2012.
32. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160*, 3–24.
33. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.
34. Minsky, M.; Papert, S. *Perceptron: An Introduction to Computational Geometry*; MIT Press: Cambridge, UK, 1969; Volume 19, p. 2.
35. Baldi, P.; Sadowski, P.J. Understanding dropout. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 2814–2822.
36. Delling, F.N.; Rong, J.; Larson, M.G.; Lehman, B.; Fuller, D.; Osypiuk, E.; Stantchev, P.; Hackman, B.; Manning, W.J.; Benjamin, E.J.; et al. Evolution of mitral valve prolapse: Insights from the Framingham Heart Study. *Circulation* **2016**, *133*, 1688–1695. [[CrossRef](#)] [[PubMed](#)]