RETRACTED: A Novel Deep Learning CNN for Heart Valve Disease Classification Using Valve Sound Detection

Randa I. Aljohani, Hanan A. Hosni Mahmoud, Alaaeldin Hafez and Magdy Bayoumi

Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdullah University, Riyadh 11671, Saudi Arabia
Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia
Department of Computer Engineering, College of Computer Science, University of Louisiana at Lafayette, Lafayette, LA 70504, USA

Correspondence: hahosni@pnu.edu.sa

Abstract: Valve sounds are mostly a result of heart valves opening and closing. Laminar blood flow is interrupted and abruptly transforms into turbulent flow, causing some sounds, and is explained by improper valve operation. It has been feasible to demonstrate that the typical and compulsive instances are different for both chronological and spatial aspects through the examination of phono-cardiographic signals. The current work presents the development and application of deep convolutional neural networks for the binary and multiclass categorization of multiple prevalent valve diseases and typical valve sounds. Three alternative methods were taken into consideration for feature extraction: mel-frequency cepstral coefficients and discrete wavelet transform. The precision of both models accomplished F1 scores of more than 98.2% and specificities of more than 98.5%, which reflects the instances that can be wrongly classified as regular. These experimental results prove the proposed model as a highly accurate assisted diagnosis model.

Keywords: deep learning; heart disease; sound classification

1. Introduction

According to the World Health Organization (WHO) [1], cardiovascular illnesses are the leading cause of death worldwide. These statistics include heart valve diseases (HVD), where moderate-to-severe valve irregularities are very typical in adults and become more prevalent as people become older [1–3].

The most frequent procedure is the auscultating process, which involves listening to acoustic features through the chest wall to assess the health of the heart valves. These cardiac acoustics can be understood as the sound expression of the tricuspid, mitral, pulmonary, and aortic heart valves opening and closing, where a pressure difference results from blood flow’s rapid acceleration and delay caused by the muscle contraction that moves blood from one cavity to another [3,4]. Its unidirectional, regular physiological operation enables proper blood flow through the cardiovascular circuit. However, some sounds are caused when laminar blood flow is interrupted by turbulent blood flow, which is explained by defective and diseased heart valve function.

Systole is the phase of the heartbeat when the ventricles of the heart push blood toward the arteries; when the ventricles fill with blood, this phase is called diastole. These two stages make up the cardiac cycle. The mitral and tricuspid valves close to start systole, which results in the first heart sound, or S1, while the aortic and pulmonic valves close to start diastole, which results in the second heart sound, or S2. During the cardiac cycle, other sounds may exist, which may point to an anomaly [5,6]. Depending on the valvar defect, the noise’s duration varies [7] as depicted in Figure 1.
Cardiac acoustics and sounds from several phonocardiographic recordings. The average time between S1 and S2 in a normal heartbeat is between 70 and 150 ms. The duration of the entire cardiac cycle is roughly 800 ms. (S1: the sound when the mitral and tricuspid valves close to start systole; S2: the sound when the aortic and pulmonic valves close to start diastole). MVP: Mitral valve prolapse, AS: Aortic Stenosis, MR: Mitral Regurgitation, MS: Mitral Stenosis (Image reused with permission [7]).

The frequency range that heart sounds produce is close to the human ear’s lowest level of sensitivity; hence, a practicing physician needs significant training under the supervision of experienced medical professionals to make an accurate diagnosis. An alternative is to use electronic stethoscopes to record heart sounds so that the medical expert can hear them to hone their hearing. Without having to rely on patients who are available during a hospital rotation, this has been shown to be beneficial in increasing physicians’ skills [8]. In both situations, the diagnosis mostly depends on the doctor’s judgment, which is susceptible to inaccuracy.

The unique feature of the current study is the use of a deep learning system to distinguish between healthy and unhealthy cardiac states by taking use of frequency dynamics that occur throughout the cardiac cycle. Three characteristics in particular mark this work as a novel methodological suggestion:

1. It has never been reported that a deep learning model and discrete wavelet transform (DWT) can transform a temporal prediction into a spatial classification problem for HVD classification because all related works use either raw time series data or feature arrangements that resemble vectors;
2. We convert a deep learning model that is pre-trained for multiclass classification into binary;
3. There are two primary stages to the deep learning algorithm’s full model. One of the distinct properties of patient cardiac cycles is processed by each of the first three parallel neural networks. The second phase combines the results of the first phase, uncovering novel features, increasing the efficacy in comparison to earlier studies.

Background

This research presents an intelligent model for the prediction of heart conditions using spatial characteristics in order to reduce the need for the practicing physician’s proper training and to produce models that can be employed as additional tools in the detection of valve diseases. The used dataset consists of phonocardiographic (PCG) records that have been classified into various cases [9–16].

The authors in [17] performed segmentation and classification of heart valve PCG signal sounds by using clustering techniques. The authors in [18] performed heart valve sound analysis and its categorization utilizing a Fuzzy Inference Model. The simulation was performed on a benchmark heart sound dataset. It was not a real-time model and was not verified with the medical subject. The authors in [19] conducted research on heart valve sounds, improving existing classification models and utilizing feature map computation of phonocardiography waves.
The authors in [20] investigated the classification of heart valve signals with deep learning modeling. They achieved an accuracy of 91.7%, but with high recall. The authors in [21] performed analysis of heart sound signals by computing the time span and the energy of valve sounds, but they only classified normal versus non-normal heart sounds. The authors in [22] performed phono-cardiographic sound analysis utilizing deep learning for abnormal heart sound prediction. Furthermore, it was utilized to differentiate normal and abnormal heart sounds. The authors in [23] and the authors in [24] performed research on PCG heart waves, utilizing discrete wavelet transform. It has serious limitations in real-time classification. The authors in [25] performed a model for heart sound prediction and detection utilizing the Kalman filter, where different feature selection algorithms were investigated. The authors in [26] proposed research on the detection of heart sounds by utilizing the Xception technique and attained good accuracy. The authors in [27] investigated heartbeat analysis utilizing the CNN-LSTM model. Thus, for PCG signal analysis, the involvement of data science and artificial intelligence is necessary.

The important thing is to recognize that similar categorization issues have already been addressed using the dataset previously described. First, the database's creators proposed performing multiclass classification [12–14]. Table 1 provides an overview of all similar works.

**Table 1.** Comparison between models that utilized the same dataset.

<table>
<thead>
<tr>
<th>Feature Selection Method</th>
<th>Model</th>
<th>Classifier</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelet transform</td>
<td>Support vector machine</td>
<td>Multiclass classification</td>
<td>88.2%</td>
<td>88.7%</td>
<td>88.4%</td>
<td>87.8%</td>
</tr>
<tr>
<td>Moments from the Frequency distribution</td>
<td>Decision tree</td>
<td>Binary</td>
<td>88.28%</td>
<td>–</td>
<td>100%</td>
<td>84.8%</td>
</tr>
<tr>
<td>Ten moments from Wavelet Transform</td>
<td>Forest tree</td>
<td>Multiclass classification</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>85.13</td>
</tr>
<tr>
<td>Frequency Approximation Entropy from discrete Fourier transform</td>
<td>C-means</td>
<td>Binary</td>
<td>88.76%</td>
<td>–</td>
<td>88.83%</td>
<td>86.5%</td>
</tr>
<tr>
<td>Entropy from discrete Fourier transform</td>
<td>multiple classifiers</td>
<td>Multiclass classification</td>
<td>88.1%</td>
<td>–</td>
<td>88.3%</td>
<td>88.33%</td>
</tr>
<tr>
<td>Raw acoustic data</td>
<td>DenseNet</td>
<td>Multiclass classification</td>
<td>–</td>
<td>–</td>
<td>88.25%</td>
<td>87.0%</td>
</tr>
<tr>
<td>Proposed model</td>
<td>DWT</td>
<td>Deep learning</td>
<td>99.5%</td>
<td>–</td>
<td>99%</td>
<td>99.2%</td>
</tr>
</tbody>
</table>

2. Materials and Methods

There are various steps involved in classifying HVDs using PCG signal analysis, as indicated in Figure 2: (i) creating the dataset and labeling it; (ii) cleaning the dataset, filtering the signal, and segmenting it according to time frames; (iii) feature selection; (iv) classifying data using a deep learning model; and (v) validation.

An issue arising from pre-processing is addressed in the current work. The signals were from an available dataset [9] with 200 entries for each class. These were converted to digital form using an 8 kHz sampling rate, with each record lasting at least one second. Each record was divided into 7000 data points (0.88 s) to ensure consistency in the data throughout the analysis. These windows must each contain at least one full cardiac cycle. Figure 3 displays the dataset’s specifics following the segmentation procedure. We present the dataset statistics in detail.
An issue arising from preprocessing is addressed in the current work. The signals were from an available dataset [9] with 200 entries for each class. These were converted to digital form using an 8 kHz sampling rate, with each record lasting at least one second. Each record was divided into 7000 data points (0.88 s) to ensure consistency in the data throughout the analysis. These windows must each contain at least one full cardiac cycle.

Figure 2. Proposed model framework.

Figure 3. Distribution of the dataset.

2.1. Feature Extraction Algorithms

The application of methods aiming at extracting informative and non-redundant parameters from the measured data is known as feature extraction in the fields of computer science, artificial intelligence, and machine learning. As a result, the stage of learning and generalization is facilitated.

We chose to employ three separate models to select the spectrum aspects of the waves: DWT and entropy.

2.1.1. Mel-Frequency Cepstral Coefficients

This research describes each step involved in making the MFCCs, but we are driven by ideas that take heart sound rates into account. A synopsis of these phases is displayed in Figure 4 [20–24].

Time series: The main goal of this phase is to split the signal into N regions, each of which is divided into a segmentation window with a shift of m and an adjacent frame separation that does not overlap. After that, each segment is subjected to a discrete Fourier transform in order to examine the frequency variations. Speech is typically processed using 20–40 ms frames with a 50% overlap (10%). However, heart sounds predominate at frequencies lower than those in speech; hence, it is advised to utilize non-overlapping 60 ms frames.

The resolution is defined by the magnitude, where N is sample size, and f is the sampling frequency. Filter bank: Despite being a hyperparameter, the number of filters' central frequencies are all determined linearly from the signal's theoretical maximum frequency. The Nyquist theorem states that the first step is to convert the scale for the signal's theoretical maximum frequency from hertz to mels [23–25].

As previously noted, this procedure of repeatedly multiplying the coefficients after applying discrete Fourier transform results in an array with N windows and M filters, as illustrated in Figure 4.
2.1. Feature Extraction Algorithms

The application of methods aiming at extracting informative and non-redundant parameters from the measured data is known as feature extraction in the fields of computer science, artificial intelligence, and machine learning. As a result, the stage of learning and generalization is facilitated.

We chose to employ three separate models to select the spectrum aspects of the waves: DWT and entropy.

2.1.1. Mel-Frequency Cepstral Coefficients

This research describes each step involved in making the MFCCs, but we are driven by ideas that take heart sound rates into account. A synopsis of these phases is displayed in Figure 4 [20–24].

![Figure 4](image-url)

**Figure 4.** Results using DWT: displaying the average value of the MFFCs for each class set on a colorimetric scale. A normalized PCG record that was randomly chosen from each sample (Reused with permission [9]).

Time series: The main goal of this phase is to split the signal into \( N \) regions, each of which is divided into a segmentation window with a shift of \( m \) and an adjacent frame separation that does not overlap. After that, each segment is subjected to a discrete Fourier transform in order to examine the frequency variations. Speech is typically processed using 20–40 ms frames with a 50% overlap (10%). However, heart sounds predominate at frequencies lower than those in speech; hence, it is advised to utilize non-overlapping 60 ms frames.

The resolution is defined by the magnitude, where \( N \) is sample size, and \( f \) is the sampling frequency. Filter bank: Despite being a hyperparameter, the number of filters’ central frequencies are all determined linearly from the signal’s theoretical maximum frequency. The Nyquist theorem states that the first step is to convert the scale for the signal’s theoretical maximum frequency from hertz to mels [23–25].

As previously noted, this procedure of repeatedly multiplying the coefficients after applying discrete Fourier transform results in an array with \( N \) windows and \( M \) filters, as illustrated in Figure 4.
2.1.2. Discrete Wavelet Transform

The discrete wavelet transform (DWT) variants represent a broad spectrum, similar to that of the short-time Fourier transform (STFT). The STFT and WT can be compared to better comprehend the WT. The WT decomposes the signal into main wavelets with various amplitudes and displacements, whereas the FT divides the signal into sines and cosines in an alternating fashion. The main wavelets have a duration that is actually constrained and an average value of 0. We used a Morlet wavelet, which is characterized as the main wavelet in the DWT, as depicted in Figure 5 [26].

![Figure 5. DWT outcomes display the averaged value of the coefficients for each class set with 150 scales, for each scale at each time point, on a colorimetric scale (reused with permission [27]).](image)

Hilbert transform: It is well-known that, when applied to the signal a priori, the Hilbert transform can enhance the DWT’s multiresolution framework [27–29].

This modification was incorporated into the model as illustrated in Figure 6 because it was demonstrated to enhance the accuracy of the deep learning network utilized.

The DWT discretizes the wavelets and captures time and scale data. Thus, similar to the DWT, a and b are the scaling and translation parameters of the parent wavelet, respectively. It is important to enter the scaling operation as specified by [29] in order to examine the spectral resolution of the data. Wavelet coefficients are what they are called. A transformation matrix made up of these coefficients is functional for the data map. We employ both low-pass filters, for fades of the signal, and high-pass filters, which only display the information’s finer details.

Mallat tree decomposition is the term for this idea of signal analysis employing filter banks. In Figure 6, we can see how the filters divide a signal into approximations and details inside the green box. The entropy and the final estimate are computed following the use of multiresolution analysis.

The wave was divided into 10 frames of 256 points, similar to the procedure described for MFCCs, and the entropy was determined by the degree of breakdown. Figure 6 displays each segment’s mean and standard deviation for each category.
This modification was incorporated into the model as illustrated in Figure 6 because it was demonstrated to enhance the accuracy of the deep learning network utilized. The DWT discretizes the wavelets and captures time and scale data. Thus, similar to the DWT, $a$ and $b$ are the scaling and translation parameters of the parent wavelet, respectively. It is important to enter the scaling operation as specified by [29] in order to examine the spectral resolution of the data. Wavelet coefficients are what they are called. A transformation matrix made up of these coefficients is functional for the data map. We employ both low-pass filters, for fades of the signal, and high-pass filters, which only display the information’s finer details.

Mallat tree decomposition is the term for this idea of signal analysis employing filter banks. In Figure 6, we can see how the filters divide a signal into approximations and details inside the green box. The entropy and the final estimate are computed following the use of multiresolution analysis.

The wave was divided into 10 frames of 256 points, similar to the procedure described for MFCCs, and the entropy was determined by the degree of breakdown. Figure 6 displays each segment’s mean and standard deviation for each category.

Figure 6. Computation of mean entropy and energy. The levels of decomposition are divided by vertical lines (reused from the public dataset no copyright is required).

2.2. Prediction

Learning is regarded as supervised learning when examples are provided with known labels (the associated accurate outputs), as opposed to unsupervised learning, where examples are unlabeled. The goal of these unsupervised (clustering) methods is to identify previously unidentified but valuable classes of items [30–32].

2.2.1. Deep Learning Model

The two primary steps of the DL used in this study are depicted in Figure 7. Three parallel artificial neural networks make up the first stage, which aims to generalize patterns. One of the networks is a multilayer perceptron, which received as input coefficients from the entropy computation following spatial factorization using discrete wavelet transform, and two CNNs, which received matrices of DWT coefficients. The outputs from the three separate networks were combined in the second stage as input to a multilayer classifier.

The final layer’s neuron count, which is two for binary classification and five for multiclass classification, determines whether the classification is multiclass or binary. However, only the multiclass network was trained, and a strategy was used to carry out the binary classification. This suggests that only the final layer was altered after the multiclass model was trained. For the multiclass classification, the anomalous class labels are divided into four subclasses, and the ReLu of the final layer varies depending on the issue. For multiclass classification, a probabilistic ReLu is used. The activation function for binary classification is a sigmoid function. The model parameters are depicted in Table 2.
Electronics 2023, 12, x FOR PEER REVIEW 8 of 14

Three separate networks were combined in the second stage as input to a multilayer classifier.

Figure 7. Block diagram for DNN: This image displays the suggested architecture. The dataset is divided into training and test sets, as shown in (A). Techniques for feature extraction (B). Algorithms for classification (C).

Table 2. The representation of the proposed platform.

<table>
<thead>
<tr>
<th>Layer #</th>
<th>Layer Type</th>
<th>Filter Dimension</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>One-layered (Input)</td>
<td>1 (200 × 200 × 200)</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>Eight layers (first six Convolutional layers)</td>
<td>6/9 × 9 × 3</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>First layer of Pooling</td>
<td>5 × 5 × 5 (max)</td>
<td>ReLU Activation</td>
</tr>
<tr>
<td>4</td>
<td>Second Convolutional layer (16)</td>
<td>16/5 × 5 × 3</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>Second layer of Pooling</td>
<td>2 × 2 × 2 (average)</td>
<td>ReLU Activation</td>
</tr>
<tr>
<td>6</td>
<td>First Dropout layer</td>
<td>0.7</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>ReLu Normalization Function</td>
<td>73</td>
<td>ReLU Activation</td>
</tr>
<tr>
<td>10</td>
<td>Second Dropout layer</td>
<td>0.3</td>
<td>–</td>
</tr>
<tr>
<td>12</td>
<td>Last layer Output</td>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>

Python 3.9 was used to implement the feature selection techniques, the deep learning model, and its components. More precisely, the DL neural network was created using the Ubuntu 20.04 distribution using Keras 2.4.3. Python-based Keras is a high-level neural network library that may be used with TensorFlow or Theano. An Intel i5-9500 processor with 64 GB of RAM was used in the model and its corresponding sub-architectures.

2.2.2. Multilayer Perceptron

A fully connected artificial neural network called a perceptron [30–34] has more than one hidden layer, and it transmits the result via an activation function. In this matrix notation, stands for the activation function, the synaptic weights of the neurons, and the output of the i neuron in the lth layer. The training synaptic weights W were updated using the descending gradient optimization algorithm [34].

3. Results

The experimentation involved creating independent classifiers during the classifier creation process that took the dataset’s features as input. To identify the most competitive configuration, each constitutive network was put to the test, both individually and in pairs.
As was previously indicated, the DL networks are first trained to achieve multiclass classification, and afterwards, a pre-trained strategy was used to change the final layer.

Figure 8 depicts the inclination for accuracy attained by the best CNN, for a subset of the dataset partitioned into a training subset and a testing subset, ranging from 60–40% to using the whole set for training. The boxplot in Figure 8 shows the F1 score accomplished by using all multiple classes more than 20 times.

![Figure 8. Accuracy of the proposed technique for various training subsets.](image)

Figure 8. Accuracy of the proposed technique for various training subsets.

Neglecting the model’s high precision obtained by employing the complete dataset for training, it can be seen that precision for the multiclass classification displays a rise resembling a quadratic curve, despite the training data percentage increasing. When training with 80% of the data, the binary classification’s performance reaches a plateau. As a result, an 80–20% split of the dataset was used for training and for testing the full model as well as its submodels.

According to precision, recall, F1 score, specificity for multiclass classification, and accuracy for binary classification, the performance of the independent models is summarized in Figure 9. With a consistent chance of deactivation of 30% for each neuron, intermediary dropout [35,36] layers were included in each network to prevent overfitting.

![Figure 9. Precision, recall, F1 score, specificity, and accuracy of each proposed model.](image)

Figure 9. Precision, recall, F1 score, specificity, and accuracy of each proposed model.
Figure 10A,C displays the performance metrics obtained for different classes in the entire model. It is easy to see that 99% of the classes in binary prediction and 95% in multiclass classification (using F1 scores) were correctly classified. Figure 10B,D also depicts the confusion matrix, which served as the foundation for all metric computations. Figure 11 depicts the confusion matrices for multiple and binary classes.

Figure 10. 20% of the dataset was used as the test set for deep learning (Precision, recall, F1 score, and specificity) of the full model in both classification techniques. To make the findings easier to see, only the percentages from 0.8 to 1.0 are displayed on the vertical axis, as shown in (A,C). Confusion matrices for binary and multiclass classification are shown in (B,D), respectively.
4. Conclusions

PCG signal analysis enables accurate signal classification that can be employed as a supplemental tool for the rapid and accurate diagnosis of HVDs. The timing, relative strength, and spectrum encoding of the PCG signals’ information enables us to apply feature selection approaches to boost their influences and speed up prediction. We made the decision to construct a deep learning model. The experiment evaluated the performance of each component of the network for a multiclass and binary classification problem both individually and in pairs. The configuration that is globally computed from the features retrieved by the DWT and the CNN model made up of multiple concurrent CNNs had the highest performance. However, the entire model’s F1 scores and binary accuracy attained values just over 95% and 99%, respectively.

Author Contributions: Methodology, H.A.H.M.; Validation, R.I.A.; Formal analysis, A.H.; Resources, M.B. Authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was funded by the Deanship of Scientific Research, Princess Nourah bint Abdulrahman University, through the Program of Research Project Funding After Publication, grant No (43-PRFA-P-16).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References


Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.