ISSN 2007-9737

Convolutional Neural Network for Improvement of Heart Valve Disease Detection

Blanca Tovar-Corona[2], Santiago Isaac Flores-Alonso[1], René Luna-García[1]

[1]Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

[2]Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
Mexico

{bltovar, rlunag}ipn.mx
sfloresa2010@alumno.ipn.mx

Abstract. Heart Valve Disease (HVD) encompasses a number of common cardiovascular conditions that account for a significant percentage of heart diseases. At present, the acoustic phenomena generated by the abnormal functioning of the heart valves can be recorded and digitized using electronic stethoscopes known as phonocardiographs. The analysis of the phonocardiographic signals has made it possible to indicate that the normal and pathological records differ in terms of both temporal and spectral characteristics. The present work describes the construction and implementation of a Deep Learning (DL) algorithm for the binary classification of normal and abnormal heart sounds. The performance of this approach reached an accuracy higher than 98 % and specificities in the "Normal" class of up to 99 %.

Keywords. Artificial intelligence, deep neural network, phonocardiography, heart valve disease.

1 Introduction

Heart noises are the expression of the opening and closing of the four cardiac valves, 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 differential [12, 15]. Its normal physiological functioning is unidirectional, which allows the correct circulation of blood through the cardiovascular circuit. However, abnormal noises can be produced when the heart valves do not close or open completely, causing leaking backwards and the interruption of laminar blood flow by turning into a turbulent flow. These sounds are called murmurs, and their correct identification during auscultation, as part of the diagnosis procedure, is crucial to detect potentially life-threatening heart conditions.

Apart from traditional auscultation, these sounds can be recorded and digitized using electronic stethoscopes, which generate phonocardiographic (PCG) signals. The identification of abnormalities of the mechanical functioning of the heart is based on a series of features extracted from the PCG recordings, where computer-aided analysis allows to identify between normal and abnormal records, since these vary among themselves with respect to their temporal and spectral characteristics.

Therefore, the precise feature extraction is key for a correct classification of heart sounds and can play an important role in assisting the medical community in speeding up and improving the diagnosis.

This article addresses the problem of identifying abnormal heart conditions using features from the PCG recording in both time and spectral domain, extracted through a technique known

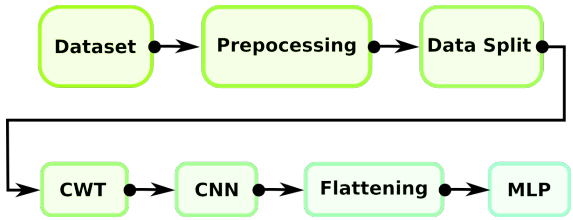

Fig. 1. Block diagram of the complete process: Dataset pre-processing and splitting, feature extraction, classification using a deep neural network

as Continuous Wavelet Transform, and a based Deep Learning (DL) methodology known as Convolutional Neural Networks (CNN). The output of the network grants the probability that a particular PCG recording belongs to a normal or abnormal class. The summary of the proposed model as a block diagram is shown in Fig. 1.

2 Related Work

Several laboratories, using particular datasets, have approached the heart sounds classification problem using their own distinctive AI methodology [14]. However, to make a correct comparison, it is necessary to select those works that use the same database as in the present work. Table 1 summarizes the feature extraction techniques and classifiers used, along with the results respectively obtained, using the same open access dataset [17].

However, a point noted is the trifle with which they approach the training of their models. It is possible to observe in [17, 18, 13] that the reported results are those obtained during training since the number of the samples shown in the confusion matrices, sums as the total of samples in the dataset.

This has important implications for the interpretation of the reported results since through training it is only possible to know the memorization capacity of the classification algorithm and the degree of compaction of the data. It is not possible to evaluate an actual performance if it is not through a test data set that the classifier has never seen.

Furthermore, the use of Convolutional Neural Networks (CNN), along with the spectral decomposition known as Continuous Wavelet Transform (CWT), has never been used to classify heart valve disease, placing the present work as a new methodological proposal.

3 Materials and Methods

This section summarizes the feature extraction techniques and Deep Learning algorithm used to address the problem apropos the HVD detection, along with the dataset description. The algorithmic proposal was developed in Python 3.9 on the Ubuntu 20.04 distribution. In particular, the deep learning algorithm was built on Keras 2.4.3.

3.1 Dataset

The PCG signals used in this article were obtained from an open database [17][1], containing 200 records for each of the following five classes:

- Aortic stenosis (AS).
- Mitral regurgitation (MR).
- Mitral stenosis (MS).
- Mitral valve prolapse (MVP).
- Normal (N).

Each signal was sampled at 8000 Hz, with durations of at least one second. To maintain uniformity in the data analysis, two windows of 6144 data points (0.768 s) were taken from each signal, each one containing at least one complete cardiac cycle, therefore, duplicating the number of samples from 200 to 400 for each class.

It is possible to notice that the Normal (N) and Pathological (AS, MR, MS, MV) classes, with a ratio of 4:1, are strongly unbalanced. This has implications for the model training, as mentioned in the previous section. Since the Normal class was separated into training and test subsets containing 320 (80%) and 80 (20%) time series, respectively, it was necessary to select the same subsets of the Pathological class to avoid the related bias. Therefore, 80 random samples of each subclass

[1] https://github.com/yaseen21khan/

Table 1. Comparative table between works that used the same dataset

Author	Feature Extraction	Classifier	Precision	Recall	Specificity	Global Accuracy
Son et al. 2018 [17]	DWT and MFCCs	SVM, KNN, DNN	–	98.2%	99.4%	97.9%
Alqudah, A. M. 2019 [4]	Eight statistical moments from the Instantaneous Frequency Estimation + PCA	KNN* and Random Forest	100%	98.28%	100%	94.8%
Ghosh, S.K. et al. 2019 [7]	Wavelet Synchrosqueezing Transform	Random Forest	–	–	–	95.13%
Upretee, P., and Yuksel, M. E. 2019 [18]	Centroid Frequency Estimation	SVM and KNN*	99.6%	99.76%	98.83%	96.5%
Ghosh, S.K. et al. 2020 [6]	Local energy and entropy from Chirplet Transform	WaveNet	98.0%	98.1%	99.3%	98.33%

(AS, MR, MS, MVP) were selected to structure the other half of the training subset.

Afterward, each time series was transformed using CWT. The implications of using this extraction technique and the procedure are discussed forward.

3.2 Continuous Wavelet Transform

CWT is a spectral decomposition method which is based on representing the signal in the form of wavelets with different displacement and scaling factors, where the use of the correct mother wavelet (MW) drives the enhancement of the waveforms of interest.

The MW is an effectively limited waveform in duration, with an average equal to zero. The MW used in the CWT was a Morlet, described by:

$$\psi(t) = e^{-\pi t^2} e^{i\pi t}. \tag{1}$$

And starting with an MW ψ, the family $\psi_{\tau,s}$ of "daughters wavelets" can be obtained by simply scaling and moving ψ:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\psi(\frac{t-\tau}{s}), \quad s, \tau \in \mathbb{R}, s \neq 0, \tag{2}$$

where s is a scaling or dilation factor that controls the width of the wavelet and τ is a translation parameter controlling its location. Scaling a wavelet simply means stretching it (if $|s| > 1$) or compressing it (if $|s| < 1$), while translating it simply means shifting its position in time [2].

Thus, the CWT of a signal $f(t)$ is given by [16]:

$$CWT(\tau, s) = \langle f, \psi_{\tau,s} \rangle \sum_{0}^{+\infty} f(t)\psi(\frac{t-s}{\tau})dt, \tag{3}$$

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

However, as a previous step to the CWT calculation, the Hilbert transform was implemented since this transform is an efficient tool to extract the time-localized amplitude and phase of a mono-component signal, with scale and translation invariance, and its energy-conserving (unitary) nature [5, 11]. The Hilbert transform $\hat{s}(t)$ of a function $s(t)$, is defined as the convolution of $(s(t) * 1/(\pi t))$ such that [9]:

$$\hat{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau. \tag{4}$$

It is possible to observe that this gives us a complex representation. To retrieve all the information of the signal, it is necessary to select a complex MW such as Morlet. By applying the

ISSN 2007-9737

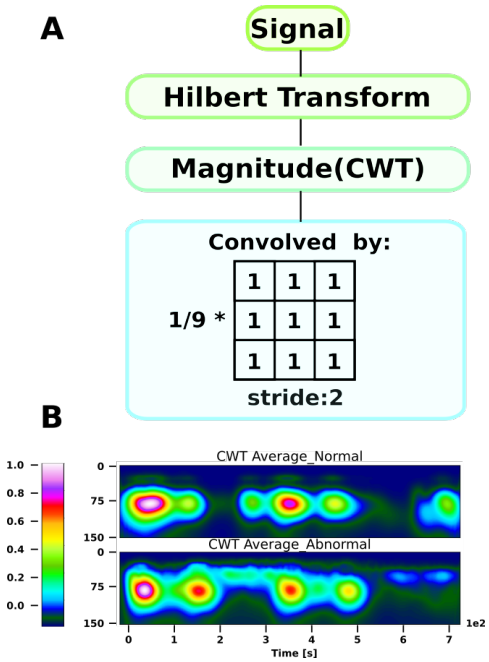

Fig. 2. CWT Results: **A)** Block diagram of the full CWT algorithm and post-processing to reduce dimensionality. **B)** Magnitude of the coefficients obtained for each scale at each time point, averaged for each of the class sets

CWT, we obtain a matrix representation of the coefficients of size NxM, as shown in Fig. 2B. To reduce computational demand, it was necessary to apply an averaging 3×3 filter as shown in Fig. 2A, which highly reduces the matrix size.

3.3 Convolutional Neural Network (CNN)

Deep learning refers to AI models capable of extracting features, with multiple levels of abstraction and learning representations of data, without the need for a human expert agent that transformed the raw data into suitable internal features from which the learning subsystem, could detect or classify patterns in the input [8].

In particular, CNN discovers intricate patterns in datasets by using the backpropagation to optimize how a set of filters need to change their internal parameters to compute the attributes that best represent the data in a highly compact depiction [10].

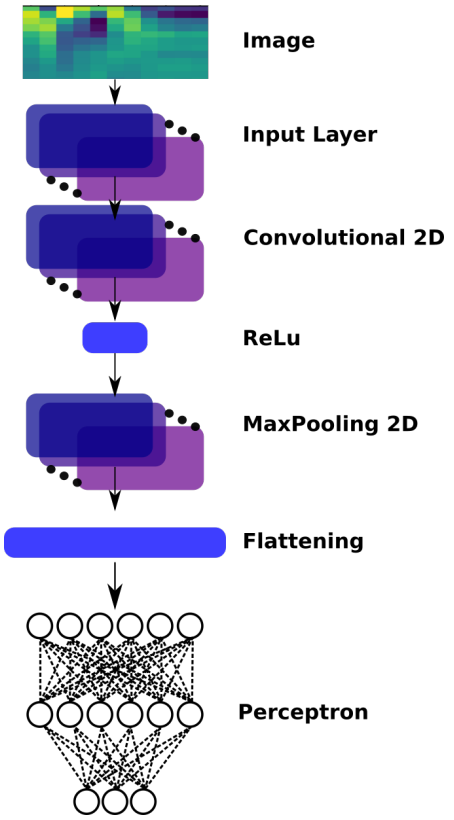

Fig. 3. Block diagram of the complete process: Dataset pre-processing and splitting, feature extraction, classification using a deep neural network

The proposition of the decomposition into a spectral space through CWT may be counterintuitive. However, since CNN uses filters that look for local spatial patterns (the locality depends on the size of the filter), the frequency dynamics of the PCG records over time contain richer information than the simple temporal dynamics of the time series.

As it is possible to see in the flow chart shown in Fig. 3, the CNN introduces a special network structure, which consists of the so-called convolution and grouping layers alternately that allow extracting the main characteristics of the coefficient matrices [1].

When using CNN for pattern recognition in phonocardiographic sounds, the input data must be organized as a series of feature maps. Since

CWT was used to find spectral coefficients along time, the expected input structure for a 2D CNN occurred naturally, where each of the coefficients represents the pixel values.

Once the input feature maps are formed, the convolution and grouping layers apply their respective operations to generate the activation of the units in those layers, in sequence. The discrete convolution between the filter and the coefficient matrix is mathematically defined as:

$$conv(I,K)_{x,y} = \sum_{i=0}^{n_{f1}-1} \sum_{j=0}^{n_{f2}-1} K_{(i,j)} I_{(x+i,y+j)}. \quad (5)$$

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

$$dim(I * K) = \Big[\frac{n_H - f1}{s} + 1, \frac{n_W - f2}{s} + 1\Big]. \quad (6)$$

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 $f1 = f2 = 2$ and $s = 2$. This operation is defined as:

$$max\big(K_{(i,j)} I_{(x+i,y+j)}\big). \quad (7)$$

In CNN terminology, the pair of convolution and max-pooling layers in succession is often referred to as a convolutional layer [3]. Each of these layers is in charge of finding, building attributes and reducing the dimensionality of the input matrix to a characteristic pattern.

Finally, this pattern is vectorized (flattened) and fed to a multilayer perceptron network (MLP), which will act as a classifier. In reality, nothing prevents the use of any other architecture or classification model, however by convention MLP is the most commonly used.

The proposed architecture of the CNN is described as pseudocode in the Algorithm 1.

3.4 Performance

The evaluation and validation of the machine learning algorithm is an essential part of any AI project. The model can give satisfactory results when it is evaluated using a metric, such as accuracy, but most of the time using a single metric is not enough to judge the performance of our model. That is why, in this section, the four evaluation metrics used are defined, where the primary building blocks are the true positive(tp), true negative(tn), false positive(fp) and false negative(fn) instantiations. In our particular case, the tp cases are the PCG recordings labelled as Normals. Therefore, the golden goal is to build a classifier with 0% fp, thus ensuring that no patient with any HDV is classified as Normal, which could pose a risk to their health and even death.

Accuracy: It is the ratio between the number of correct predictions and the total number of input samples:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}. \quad (8)$$

Due to its construction, this metric is not ideal when the classes are unbalanced (as is the case with the dataset used). The problem arises when the cost of misclassifying samples from minor classes is very high. If we are faced with a rare but fatal disease, the cost of not diagnosing a sick person's illness is far greater than the cost of sending a healthy person for further tests. Therefore, it is necessary to use metrics based on relevance, that is, that do take into account the imbalance of the classes, such as precision and recall.

Precision and recall: Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances:

$$Precision = \frac{tp}{tp + fp}. \quad (9)$$

While recall (also known as sensitivity) is the fraction of relevant instances that were retrieve:

$$Recall = \frac{tp}{tp + fn}. \quad (10)$$

Algorithm 1 CNN Architecture

1: **input:** CWT
2: $CNN \leftarrow Convolution\ layer\ (16\ Filters\ (21 \times 21))$
3: $Batch\ Normalization\ +\ Nonlinear\ layer\ +\ MaxPooling$
4: $CNN \leftarrow Convolution\ layer\ (8\ Filters\ (11 \times 11))$
5: $Batch\ Normalization\ +\ Nonlinear\ layer\ +\ MaxPooling$
6: $CNN \leftarrow Convolution\ layer\ (4\ Filters\ (7 \times 7))$
7: $CNN \leftarrow Flatten()(CNN)$
8: $MLP \leftarrow 2\ output\ neurons$
9: **output:** $Membership\ probability$

Finally, since in a clinical test the goal is to accurately identify people who have a particular condition (where its misclassification into a non-pathological class could be fatal), the ratio between true negatives and false positives should be accounted for, giving rise to a metric known as specificity. In other words, specificity measures how the test is effective when used on negative individuals:

$$Specificity = \frac{tn}{tn + fp}. \tag{11}$$

4 Results

During the construction of the model, the experimentation focused on two variables: the number of scales to be used in the CWT and the generation of the training subset, which as mentioned in section 3.1, is partially built from 320 pseudorandomly selected items from the Pathological (AS, MR, MS, MVP) subclasses.

For the case of CWT, the value of the power coefficients obtained for each scale at each time point, averaged for each of the class sets, with 150 scales is shown in Fig 2. The number of scales depends on the MW used to perform the decomposition, since each MW has a specific morphology and central frequency that will change as a function of scale. There is an approximate relationship between scale and frequency defined as:

$$s(fr) = \frac{ln(\frac{cf * fs}{fr})}{ln(2)}, \tag{12}$$

where s is the approximate scale, cf is the central frequency of the MW, fs is the sampling frequency and fr is the target frequency to approximate. However, this approximation is not exact and that is why the selection of the MW, number of scales and subscales can be defined as a hyperparameter. For the PCG records used in the present work, the 150 scales of the complex Morlet proposed as MW showed the level of detail sought.

On the other hand, to ensure that the CNN's performance was since the optimum (local) minimum was found, which ensures the generalizability of the model, and not from the pseudo-random selection of the data, a 6-fold cross-validation method was applied, where the overall accuracy obtained was 97.70 $\pm$ .432.

By having an overall accuracy with a standard deviation of less than 0.5%, the proposed model execution can be attributed to its generalizability, which allows us to select the best of the runs of our classifier to evaluate its performance. Fig. 4B shows the detailed accuracy, precision, recall and specificity obtained using 20 % of the dataset as the test set. It is possible to observe that 98.2% of the classes were correctly classified according to the binary accuracy.

Furthermore, the confusion matrix, from where all the metric calculations were based, is shown in Figure 4A, where each column of the matrix represents the number of predictions of each class, while each row represents the instances in the real class.

5 Conclusion

This article focused on the classification of HVD from 1000 PCG records, combining a deep learning algorithm with time-frequency analysis wherein the time-series recognition problem is transformed into an image recognition problem. To do so, the spectral characteristics through time were extracted using CWT, and given the dimensional nature of these features, it was decided to use a CNN to classify each recording as Normal or Pathological, since this is the first step in the diagnostic procedure. If an abnormality is present, further clinical tests must be carried out to determine the type of abnormality. This approach

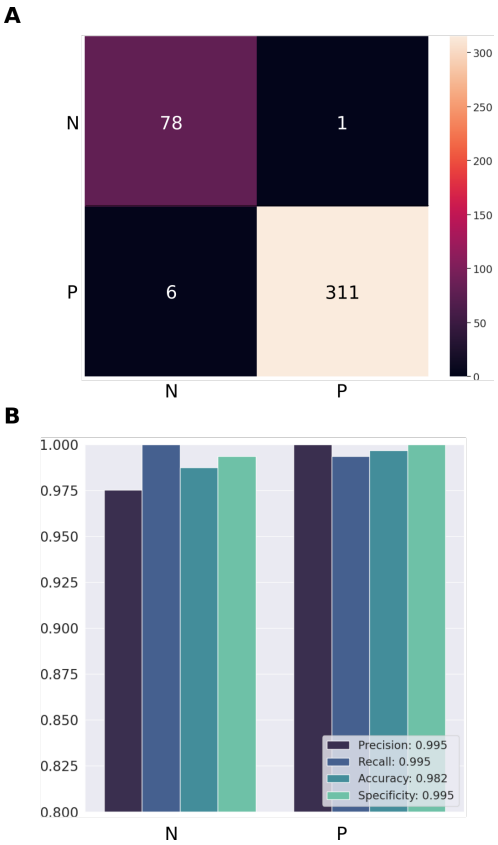

Fig. 4. DL performance results (20% of the dataset as the test set): **A** Confusion matrix used to calculate the metrics. **B** Precision, Recall, Specificity and Accuracy for the test dataset classification. The vertical axis shows only the percentage from 0.8 to 1.0 to facilitate the visualization of the results. ”N” and ”P” stand for Normal and Pathological class respectively

has never been used, placing it as an innovative methodological proposal.

Furthermore, the model had a performance, measured through its accuracy, above 98.2%, surpassing four of the five models described in the literature (Table 1), placing it as a competitive and efficient model for the classification of valvular diseases.

In addition, one of the necessary metrics to measure competitiveness in clinical diagnostic systems, and where the present work takes into account and stands out, is specificity (section 3.4), obtaining 99.5%, which means that less than 1% of the Pathological PCG records will be classified as Normal.

This provides robustness to the model and invites to implement it in a system for the assisted diagnosis of heart valve diseases to improve the prognosis of patients, reducing the error associated with the experience of the medical crew.

Acknowledgments

This research was funded by the Instituto Politécnico Nacional through the project SIP20210473. We thank CONACyT for partial support of the present work.

References

1. **Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, L., Penn, G., Yu, D. (2014).** Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing, Vol. 22, No. 10, pp. 1533–1545.
2. **Aguiar-Conraria, L., Soares, M. J. (2014).** The continuous wavelet transform: Moving beyond uni-and bivariate analysis. Journal of Economic Surveys, Vol. 28, No. 2, pp. 344–375.
3. **Albawi, S., Mohammed, T. A., Al-Zawi, S. (2017).** Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), Ieee, pp. 1–6.
4. **Alqudah, A. M. (2019).** Towards classifying non-segmented heart sound records using instantaneous frequency based features. Journal of medical engineering & technology, Vol. 43, No. 7, pp. 418–430.
5. **Chaudhury, K. N., Unser, M. (2009).** Construction of Hilbert transform pairs of wavelet bases and Gabor-like transforms. IEEE Transactions on Signal Processing, Vol. 57, No. 9, pp. 3411–3425.
6. **Ghosh, S. K., Ponnalagu, R., Tripathy, R., Acharya, U. R. (2020).** Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with pcg signals. Computers in biology and medicine, Vol. 118, pp. 103632.

7. **Ghosh, S. K., Tripathy, R. K., Ponnalagu, R., Pachori, R. B. (2019).** Automated detection of heart valve disorders from the pcg signal using time-frequency magnitude and phase features. IEEE Sensors Letters, Vol. 3, No. 12, pp. 1–4.
8. **Goodfellow, I., Bengio, Y., Courville, A. (2016).** Deep learning. MIT press.
9. **Johansson, M. (1999).** The Hilbert transform. Mathematics Master's Thesis. Växjö University, Suecia. Disponible en internet: http://w3. msi. vxu. se/exarb/mj_ex. pdf, consultado el, Vol. 19.
10. **Liu, T., Fang, S., Zhao, Y., Wang, P., Zhang, J. (2015).** Implementation of training convolutional neural networks. arXiv preprint arXiv:1506.01195.
11. **Mahato, S., Teja, M. V., Chakraborty, A. (2017).** Combined wavelet–Hilbert transform-based modal identification of road bridge using vehicular excitation. Journal of Civil Structural Health Monitoring, Vol. 7, No. 1, pp. 29–44.
12. **Mondal, A., Kumar, A. K., Bhattacharya, P., Saha, G. (2013).** Boundary estimation of cardiac events s1 and s2 based on Hilbert transform and adaptive thresholding approach. 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT), IEEE, pp. 43–47.
13. **Oh, S. L., Jahmunah, V., Ooi, C. P., Tan, R.-S., Ciaccio, E. J., Yamakawa, T., Tanabe, M., Kobayashi, M., Acharya, U. R. (2020).** Classification of heart sound signals using a novel deep wavenet model. Computer Methods and Programs in Biomedicine, Vol. 196, pp. 105604.
14. **Rajagopalan, V., Cao, H. (2022).** Cardiovascular applications of artificial intelligence in research, diagnosis, and disease management. In Biomedical and Business Applications Using Artificial Neural Networks and Machine Learning. IGI Global, pp. 80–127.
15. **Randhawa, S. K., Singh, M. (2015).** Classification of heart sound signals using multi-modal features. Procedia Computer Science, Vol. 58, pp. 165–171.
16. **Sinha, S., Routh, P. S., Anno, P. D., Castagna, J. P. (2005).** Spectral decomposition of seismic data with continuous-wavelet transform. Geophysics, Vol. 70, No. 6, pp. P19–P25.
17. **Son, G. Y., Kwon, S. (2018).** Classification of heart sound signal using multiple features. Applied Sciences, Vol. 8, No. 12, pp. 2344.
18. **Upretee, P., Yüksel, M. E. (2019).** Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT´19), IEEE, pp. 1–4.

Article received on 08/04/2022; accepted on 25/05/2022.
Corresponding author is René Luna-García.