ISSN: 1692-7257 - Volume 2 – Issue 48 - 2026

ORIGINAL

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

Assessing the robustness of information theory descriptors in cardiac signal classification

Evaluación de la robustez de los descriptores de teoría de la información en la clasificación de señales cardíacas

Lely Adriana Luengas Contreras[1]
Esperanza Camargo Casallas[1]
Jonathan Roberto Torres Castillo[1]

[1] **Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.**

Received: 23 February 2026. Revised: 09 June 2026. Accepted: 24 July 2026. Published: 28 July 2026.
Corresponding Author: Lely Adriana Luengas Contreras

How to cite: L. A. Luengas Contreras, E. Camargo Casallas, and J. R. Torres Castillo, "Evaluación de la robustez de los descriptores de teoría de la información en la clasificación de señales cardíacas", *Revista Colombiana de Tecnologías de Avanzada (RCTA)*, vol. 2, no. 48, jul. 2026, art. no. 4593, doi: 10.24054/rcta.v2i48.4593

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract:

Automated classification of electrocardiographic (ECG) signals can be a useful tool for identifying cardiovascular abnormalities. However, the nonlinear and nonstationary nature of these signals may limit the use of conventional parametric methods. This study evaluates Information Theory (IT) descriptors as a nonparametric alternative for characterizing heartbeat segments. A total of 14,552 segments from the ECG Heartbeat Categorization Dataset were analyzed, each represented by 187 temporal positions, and the interdependence between diagnostic categories and amplitudes was quantified using entropy and mutual information. An IT based clustering procedure was also implemented. For the abnormal category, the confusion matrix showed a sensitivity of 94.38% and a specificity of 70.53%; balanced accuracy was 82.46%. Cluster purity was close to 0.877. The results indicate partial separation between normal and abnormal segments and should be interpreted considering class imbalance, preprocessing, and the effect of zero-padding.

Keywords: Shannon entropy, complexity analysis, ECG signal processing, mutual information, least information lost, information theory.

Resumen:

La clasificación automatizada de señales electrocardiográficas (ECG) puede constituirse en una herramienta de apoyo para la identificación de alteraciones cardiovasculares. No obstante, la naturaleza no lineal y no estacionaria de estas señales puede limitar el uso de métodos paramétricos convencionales. Este estudio evalúa descriptores de Teoría de la Información (TI) como alternativa no paramétrica para caracterizar segmentos de latidos cardíacos. Se analizaron 14 552 segmentos del ECG Heartbeat Categorization Dataset, cada uno representado por 187 posiciones temporales, y se cuantificó la interdependencia entre las categorías diagnósticas y las amplitudes mediante entropía e información mutua. Además, se implementó un procedimiento de agrupamiento basado en TI. La matriz de confusión para el modelo de clasificación mostró una sensibilidad de 94,38 % y una especificidad de 70,53 %; la exactitud balanceada fue 82,46 %. La pureza de los clústeres fue cercana a 0,877. Los resultados muestran una separación parcial entre segmentos normales y anormales y deben interpretarse considerando el desbalance de clases, el preprocesamiento y el efecto del zero-padding.

Palabras clave: entropía de Shannon, análisis de complejidad, procesamiento de señales de ECG, información mutua, menor pérdida de información, teoría de la información.

INTRODUCTION

In recent years, the use of information theory (IT) and its descriptors, such as entropy, complexity and dependence measures (mutual information, channel capacity, etc.), in the analysis of biomedical signals in EEG, ECG, PPG, intracranial pressure and medical images has become consolidated owing to the advantages it offers: it does not depend on the type of data and its statistical distribution, but rather analyzes the amount of information that the dataset and each of the variables possess [1]-[6]. Reviews focused on IT and carried out between 2015 and 2024 have systematized its use in physiological signals, evidencing a significant expansion in the measurement of various biological parameters and their domains of application in biomedical engineering[2],[3], [7]-[10]. The use of information theory quantifiers (entropy and the complexity–entropy plane) in electrical biosignals has made it possible to characterize disorder, organization and structure in various pathologies, identifying gaps especially in digestive disorders[2].

In biological systems, Shannon's theory has been used to quantify information transmission and channel capacity in signaling pathways and to infer molecular interaction networks, highlighting problems of sample size and high dimensionality, problems that are common in biomedicine[11].

In the EEG domain, the combination of complexity quantifiers with advanced signal processing techniques and machine learning has been used for neurological diagnosis, brain–computer interfaces and clinical monitoring, ranging from the classification of stroke, wake/sleep states and lie detection to emotion recognition, with the concepts associated with IT being fundamental in these applications, showing high-quality results [12]-[15].

As for the ECG, for the classification of heartbeats and arrhythmias, spectral entropy calculated in the time–frequency domain has been used together with convolutional neural networks, achieving effective discrimination between different types of heartbeats[16],[17]. Similarly, wavelet entropy or "wavelet packet entropy" has been combined with random forest-type algorithms for arrhythmia classification, reaching results superior to other traditional approaches and demonstrating the usefulness of entropic measures as robust descriptors of ECG morphology and variability[17]. In the detection of coronary disease, it has been shown that entropy analysis applied to ECG from healthy subjects and patients with coronary artery disease presents acceptable results. Complementarily, hierarchical and multiscale entropy approaches allow the nonlinear and multifrequency characteristics of the ECG to be evaluated, integrating low- and high-frequency information and providing discrimination rates between normal records and heart failure when fused with probabilistic neural networks[18]. Recent reviews on multiscale entropy in physiological signals support these findings and underscore the role of the multiscale complexity of the ECG as a sensitive marker of cardiovascular pathological states[18]; hence, quantifiers derived from information theory can constitute a central axis in ECG processing [16]-[20].

Mutual information and transfer entropy measures have also been used to study the flow of information between the twelve standard leads, identifying which leads contribute the greatest informative content and how information is shared among them. This has direct implications for optimal lead selection and for the design of diagnostic systems with fewer channels, such as Holter monitors or portable devices [21]-[25]. Along the same lines, redundancy metrics based on mutual information have been proposed to quantify the informational overlap between leads and to optimize the input of convolutional neural networks, maximizing relevant information and reducing redundancy without degrading diagnostic performance[22].

ECG signal descriptors based on information theory, such as entropy and mutual information, have been reported in previous works; however, their integration with a discretization aimed at maximizing the amount of information in a system without assuming a normal distribution, yet allowing normal and abnormal segments to be separated, has been little explored. It is hypothesized that discretization by means of LIL preserves part of the relevant diagnostic information and allows identifying dependence between the segments of measured ECG values and the type of signal. In this scenario, the contribution of this work lies in carrying out a quantitative evaluation that uses Shannon's information theory to integrate LIL categorization, Shannon entropy, mutual information and the generation of a classification model, as well as examining the incidence of zero-padding on the descriptors obtained. The contribution therefore focuses on a methodological validation applied to ECG signals.

Information Theory

A system, regardless of its nature, can be described as an information system using mathematical expressions. In nonlinear physical systems (for example, biological systems), linear correlation is an imprecise tool. A biological system X exists in the state space with a probability distribution p(x). Its average uncertainty, or the potential for information transmission, is defined by the Shannon entropy (H) (1)[26].

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_b P(x_i) \quad (1)$$

Where $P(x_i)$ is the probability of a specific event. When b=2, the unit is the bit.

The relationship between variables in a system is similar to the transfer of information in a communication channel. Each

channel has a limit conditioned by the physical part and by noise. Shannon's theorem defines the maximum rate (C) at which data can be transmitted without error in a noisy channel (2)[26].

$$C = B\log_2\left(1 + \frac{S}{N}\right) \quad (2)$$

Where B is the bandwidth (Hertz); S/N is the signal-to-noise ratio.

The reduction of uncertainty of X due to knowledge of Y is called mutual information I(X;Y) (3)[26].

$$I(X;Y) = H(X) - H(X|Y) \quad (3)$$

In an ideal system, I(X;Y) is equal to the total entropy of the source. In practice, the noise H(X|Y) represents the equivocation or uncertainty of the input upon knowing the output (4)[26].

$$H(X|Y) = -\sum_x p(x|y)\log_2 p(x|y) \quad (4)$$

METHODS

For the analysis, the binary subset of the ECG Heartbeat Categorization Dataset, derived from the PTB Diagnostic ECG Database and publicly available[27], was used. The set contains 14 552 ECG heartbeat segments: 4 046 labeled as normal and 10 506 as abnormal. Each row includes 187 temporal amplitude values and an additional diagnostic label; therefore, the analyzed matrix contains 2 721 224 amplitude values (14 552 × 187). The 187 positions do not correspond to a selection of subjects or to a sampling of observations, but rather to the fixed length of each segment. At a sampling frequency of 125 Hz, this length is equivalent to approximately 1,496 s. All available segments from the two categories were included, and no random sampling was performed. The diagnostic label was used as the output variable and was not included among the 187 analyzed amplitudes. Because the preprocessed files are organized by segments, the counts should not be interpreted as numbers of independent patients.

The data were extracted from CSV files and organized for analysis. The methodology proposed byLuengas (2016)[26]was taken as the basis; its general sequence is presented in Figure 1. First, an exploratory data analysis was performed to characterize the behavior of the signals. Measures of central tendency and dispersion were calculated to describe the variability between segments and to compare the normal and abnormal categories.

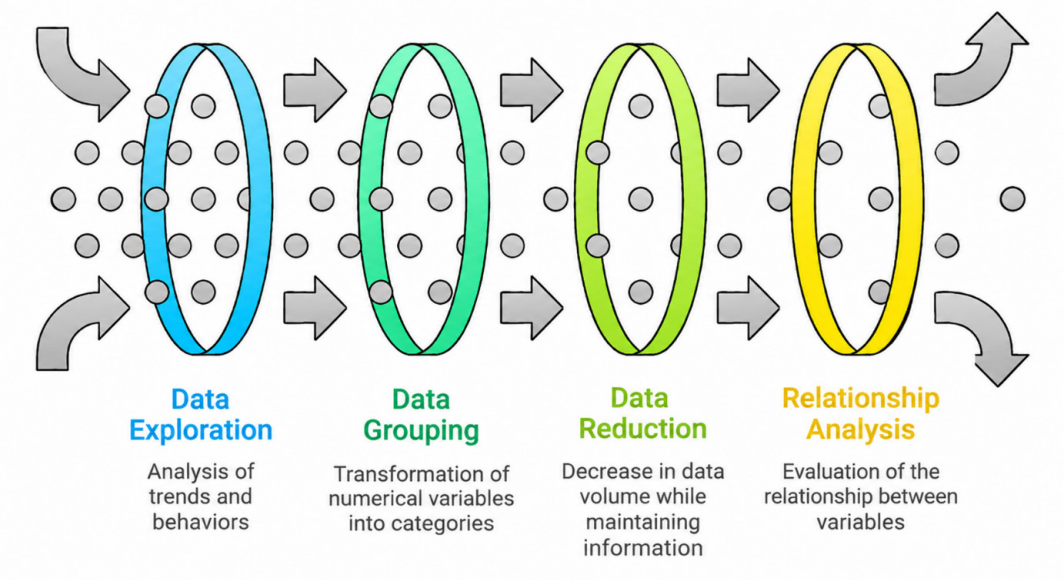

Figure 1. *Scheme of the methodology for the analysis of ECG signals.*
Source: own elaboration.

Once the data had been examined statistically, the data grouping stage was carried out. Since the foundations of Information Theory (IT) require the handling of discrete variables, and considering that ECG amplitude measurements are continuous (numerical) in nature, the discretization or grouping (binning) process called Least Information Loss (LIL) was implemented. LIL calculates the information transmitted by the variables to evaluate feasible category boundaries; the selected intervals frame the data that best describe the behavior of the system with the least increase in entropy. In this way, the intervals or

bins used in the informational analysis are obtained.

The amount of categorical data was reduced to lower the computational load and obtain a simplified representation of the system. The procedure sought to preserve the positions with the most information relative to the output. To this end, an algorithm for calculating the transmitted information and the signal-to-noise ratio (SNR) was used; data with an SNR greater than 70% were discarded, since the information they contain is low and their correspondence with the input is minimal[15].

Finally, the relationship between the input temporal positions and the output diagnostic category was analyzed using the free version of Powerhouse®. Two clusters were set, corresponding to the two categories of the set: normal and abnormal. The procedure integrated the transmitted information, the signal-to-noise indicator reported by the software, and the verification of non-collinearity between variables. The correspondence between the cluster assignment and the diagnostic labels was evaluated using a confusion matrix.

To interpret the confusion matrix, category 1 (abnormal) was treated as positive. The columns represent the expected categories and the rows the assignment obtained; the percentages are normalized by expected category. Sensitivity, specificity, precision, overall accuracy, balanced accuracy and F1 were calculated. As external clustering indicators, purity, the Rand Index (RI), the Adjusted Rand Index (ARI) and normalized mutual information (NMI) were estimated from the contingency table. Because the figure reports percentages rounded to two decimals, the reconstructed counts and the derived metrics are presented as approximate values.

Data availability

The preprocessed set used in this study corresponds to the binary subset of the ECG Heartbeat Categorization Dataset, derived from the PTB Diagnostic ECG Database and publicly available on Kaggle [27]. The analysis was performed on heartbeat segments and not on records identified at the patient level.

RESULTS

Data exploration

The temporal and statistical variability of the ECG segments classified as normal or abnormal was analyzed. The data, sampled at 125 Hz and normalized in amplitude, were subjected to a non-parametric analysis to identify potentially discriminative statistical descriptors. Figure 2 shows some of the signals analyzed.

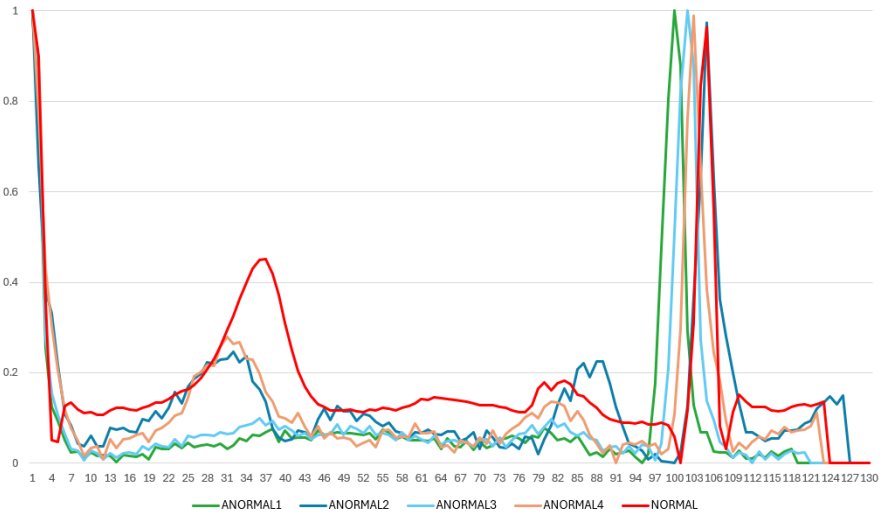

Figure 2. *Sample of the signals analyzed.*
Source: own elaboration.

Distribution and Normality

The analysis of the amplitude distribution across the 187 temporal positions of the segments made it possible to establish the analytical framework. The D'Agostino–Pearson test confirmed that the data do not follow a normal distribution ($p < 0.001$).

The comparison of amplitudes between categories was performed using the Mann–Whitney U test and showed statistically significant differences in more than 96 % of the temporal positions. The Kolmogorov–Smirnov test also indicated

differences between the distributions. The concentration of values equal to zero at the final positions is related to the zero-padding used to complete the shorter segments up to 187 positions; therefore, these positions may introduce bias into the statistical and informational measures.

Measures of tendency

Given the absence of normality, the median and the root mean square (RMS) were used as estimators of central tendency and energy, respectively. Differences in amplitude, local energy and median were observed between categories, which indicates that the ECG signals describe morphological variations in the segments, without by themselves establishing a clinical diagnosis.

Variability between segments

The analysis reveals that normal signals exhibit a highly predictable morphology with a reduced interquartile range (IQR). In contrast, the group with pathologies shows significantly greater dispersion ($p < 0.01$, Mann-Whitney U test), which suggests that the pathologies represented do not converge on a single morphology but rather increase the entropy of the system.

Energy Descriptors (RMS)

The RMS analysis revealed that subjects with cardiac alterations show higher values than subjects without alterations. This finding may be linked to abnormalities, but it cannot be attributed to a specific clinical condition without additional diagnostic information.

Summary of the Statistical Analysis

The statistical analysis revealed differences between the normal and abnormal categories. The abnormal category showed a lower median peak amplitude (0.92 ± 0.18 vs. 1.00 ± 0.04; $p < 0.05$) and a higher RMS (0.3128 vs. 0.2741; $p < 0.01$). It also showed lower kurtosis (9.8 vs. 14.2; $p < 0.001$) and a larger interquartile range (> 0.28 vs. < 0.15; $p < 0.001$). These results describe greater morphological and energetic variability in the abnormal segments. Table 1 summarizes the statistical data.

Table 1: Summary of the statistical measures

Statistical Parameter	Normal Group	Abnormal Group	Sig (p)
Median	1.00 ± 0.04	0.92 ± 0.18	< 0.05
RMS	0.2741	0.3128	< 0.01
Kurtosis	14.2	9.8	< 0.001
IQR	Low (< 0.15)	High (> 0.28)	< 0.001

Source: own elaboration

Data Clustering and Reduction

Discretization by means of LIL sought to preserve the relevant information of the amplitudes with respect to the output category and to reduce the distortion introduced during quantization. The transformation generated specific partitions for the temporal positions and allowed a reduced representation to be selected. A 56 % reduction of the temporal positions was obtained, so 44 % of the inputs with the greatest transmitted information according to the implemented criterion remained.

Analysis of Relationships

The data were analyzed by means of IT to obtain the infometric map of Figure 3. Powerhouse® reported a 44 % relationship between the information of the selected inputs and the output. This percentage evidences a partial informative association between the temporal positions of the ECG and the diagnostic category, but it does not constitute a direct measure of accuracy, nor is there a universal threshold that allows it to be qualified in isolation as high or low. The existence of a relationship indicates that the system can be modeled, since the variables are not independent[28].

In the software representation, 44 % is equivalent to 0.44 bits of transmitted information, and the remaining percentage corresponds to non-transmitted information or error within that scheme. The practical usefulness of the value must be analyzed together with the confusion matrix and the external clustering indicators.

As indicated, the infometric map supports the existence of an interdependence between the input variables and the output variable, but by itself it does not show the possibility of creating a high-performance predictive model.

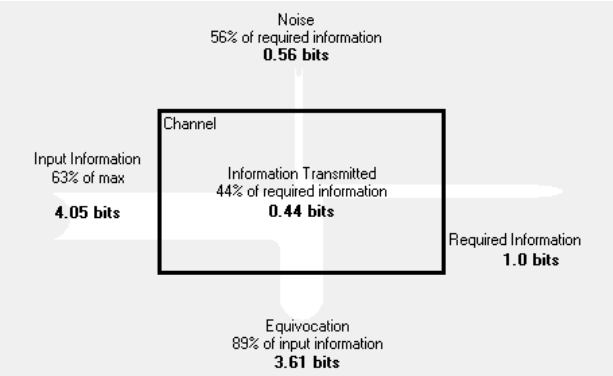

Figure 3. *Infometric map of the dataset. The relationship between the inputs and the output is 44%.*
Source: own elaboration

Clustering

With the purpose of validating the performance of the analysis based on Information Theory (IT), a classification model was implemented with the clustering technique, in which the segregation and comparison of two distinctive cohorts was contemplated: a group composed of normal ECG records and another representative of pathological cardiac morphologies. The Powerhouse® software was used.

In contrast to traditional approaches based on Euclidean distances, the intra-cluster (cohesion) and inter-cluster (separation) similarity measures were based on the transmitted mutual information, considering the previous analysis; in this way, the dependence between the input vectors (ECG samples) and the output classification of the dataset is quantified; thus, the consolidation of the clusters responds both to the geometric proximity of the signals and to the informative coherence and the reduction of the diagnostic uncertainty inherent to the system.

Powerhouse® calculated the mutual information, the untransmitted information and the equivocation, and represented the clustering by means of the density map in Figure 4. Cluster 1 showed a higher concentration of abnormal segments and cluster 2 a higher concentration of normal segments.

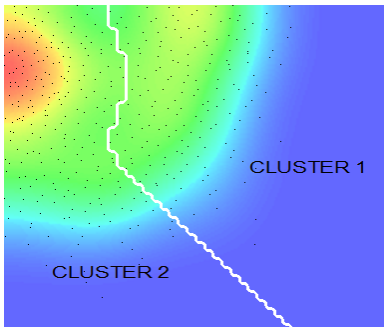

Figure 4. *Density map.*
Source: own elaboration

Figure 5 presents the distribution of the categories in the initial dataset and in each cluster. The representation shows a partial separation between normal and abnormal segments; therefore, the results suggest discriminative capacity, but they do not imply a perfect separation or a clinical validation of the procedure.

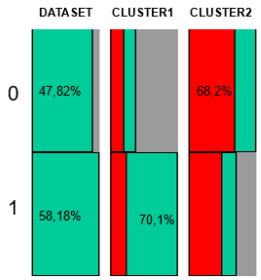

Figure 5. *Distribution of the data in each cluster*
Source: own elaboration

Figure 6 presents the normalized confusion matrix. Considering category 1 as abnormal, a sensitivity of 94.38 % and a specificity of 70.53 % were obtained. From the class sizes and the rounded percentages, an overall accuracy of 87.75 %, a precision of 89.27 %, a balanced accuracy of 82.46 % and an F1 of 91.75 % were estimated. As external clustering metrics, a purity of 0.877, a Rand Index of 0.785, an Adjusted Rand Index of 0.545 and an NMI of 0.391 were approximately obtained. These values were derived from the reconstructed contingency table and should be confirmed with the original counts if these are available.

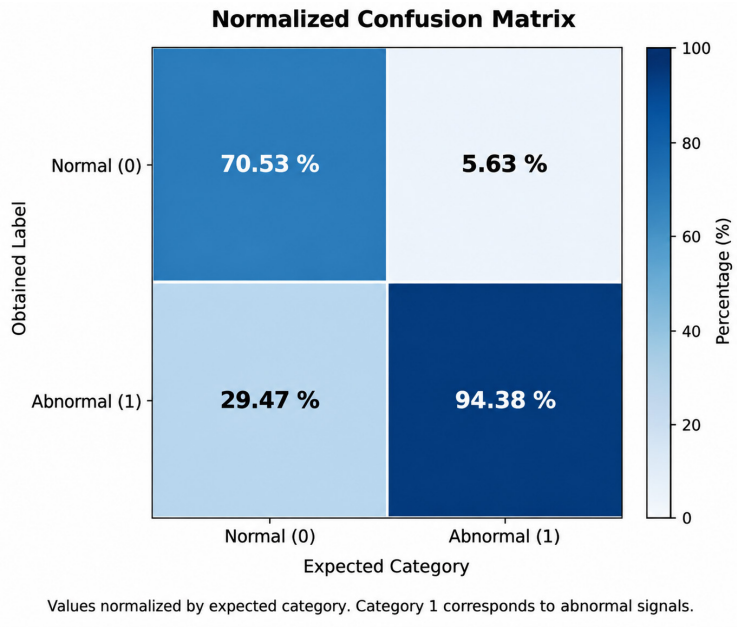

Figure 6. *Distribution of the data in each cluster*
Source: own elaboration

Discussion and limitations

Entropy and mutual information measures have been used in previous research as descriptors together with machine learning classifiers [16]–[20]. In contrast, the present work estimated the transmitted information in order to investigate the amount of information contained after the discretization of the data. Hence, the 44% of transmitted information shows a relationship between inputs and output, but not an estimate of the performance of a computational model; for this, the sensitivity, the specificity and the external metrics of the cluster model must be analyzed.

In order to have a specific format for the data, zero-padding was used; in this way, zeros are added at missing value positions; however, modification of the entropy, the mutual information and the measures of central tendency may occur.

Among the limitations of the research are the use of a single preprocessed dataset, the analysis at the level of ECG signal segments and the limited information on the algorithms used in Powerhouse®, as well as its internal configuration. Additional validation with other, unprocessed data is needed, as is strengthening the performance metrics of the model before generalizing the results.

CONCLUSIONS

There is a statistically significant difference between normal and abnormal ECGs, both in terms of amplitude and in the shape of the distribution, confirmed by non-parametric tests (Mann–Whitney U and Kolmogorov–Smirnov). These results validate the use of automatic classification techniques and justify the extraction of temporal and morphological features for cardiac anomaly detection.

Most temporal samples are significantly distinct, as shown by the statistical analysis. This suggests that the cardiac pathology does not affect only an isolated point but modifies the global structure of the ECG signal.

To group ranges of values into categories or intervals and thereby reduce the variability (eliminating part of the fine structure) in a dataset, discretization was performed using LIL. This technique made it possible to transform numerical data into categorical data while preserving the transmitted information. A transmitted information of 44% is reported, referring to the relationship between inputs and output, but it is not a measure of model performance.

The performance of the model with the confusion matrix shows a balanced accuracy for a model that allows discrimination between data, but that must be validated with another dataset.

Zero-padding especially influences the last positions of the segments, affecting the statistical measures and those associated with the TI; hence, the results linked to those positions must be evaluated with caution.

Due to the non-parametric nature of the data, any classification model (such as neural networks or support vector machines) must be trained considering that the arithmetic mean is not a reliable descriptor; the use of the Median and the RMS is suggested.

In general, the results support the use of descriptors associated with the TI for ECG discrimination. However, independent validation is required before establishing their usefulness as the basis of a predictive model or extrapolating the findings to other clinical contexts.

REFERENCES

[1] K. Balasubramanian, S. Rajendran, and S. Pati, "Complexity Measures in Biomedical Signal Analysis: A Clinically-Grounded Survey Across EEG, ECG, Intracranial Pressure, and Photoplethysmogram Modalities," *IEEE Access*, vol. 13, pp. 155285–155304, 2025. doi: 10.1109/access.2025.3603848

[2] L. Acosta and M. Pacheco, "Entropy and Statistical Complexity in Bioelectrical Signals: A Literature Review," *Signals*, vol. 6, no. 1, p. 7, 2025. doi: 10.3390/signals6010007

[3] J. Li, G. Feng, C. Ling, X. Ren, X. Liu, S. Zhang, L. Wang, Y. Chen, X. Zeng, and R. Chen, "A Resource-Efficient Multi-Entropy Fusion Method and Its Application for EEG-Based Emotion Recognition," *Entropy*, vol. 27, no. 1, p. 96, 2025. doi: 10.3390/e27010096

[4] C. Wu, B. Yao, X. Zhang, T. Li, J. Wang, and J. Pu, "The Application of Entropy in Motor Imagery Paradigms of Brain–Computer Interfaces," *Brain Sci.*, vol. 15, no. 2, p. 168, 2025. doi: 10.3390/brainsci15020168

[5] D. Platakis and G. Manis, "Review of Recent (2015–2024) Popular Entropy Definitions Applied to Physiological Signals," *Entropy*, vol. 27, no. 9, p. 983, 2025. doi: 10.3390/e27090983

[6] L. A. Luengas and L. F. Wanumen Silva, "Modelos computacionales en la posturografía," *Tecnura*, vol. 26, no. 73, pp. 30–48, 2022. doi: 10.14483/22487638.18060.

[7] Zandbagleh, S. Sanei, L. Penalba-Sánchez, P. Rodrigues, M. Crook-Rumsey, and H. Azami, "Intra- and Inter-Regional Complexity in Multi-Channel Awake EEG Through Multivariate Multiscale Dispersion Entropy for Assessing Sleep Quality and Aging," *Biosensors*, vol. 15, no. 4, p. 240, 2025. doi: 10.3390/bios15040240

[8] L. Sparacino, Y. Antonacci, G. Mijatović, and L. Faes, "Measuring hierarchically-organized interactions in dynamic networks through spectral entropy rates: Theory, estimation, and illustrative application to physiological networks," *Neurocomputing*, vol. 630, p. 129675, 2025. doi: 10.1016/j.neucom.2025.129675

[9] D. Sengupta, P. Gupta, and A. Biswas, "An efficient similarity metric for 3D medical image registration," *Multimedia Tools Appl.*, vol. 83, pp. 87987–88017, 2024. doi: 10.1007/s11042-024-18710-1

[10] M. Zbili and S. Rama, "A Quick and Easy Way to Estimate Entropy and Mutual Information for Neuroscience," *Front. Neuroinform.*, vol. 15, p. 596443, 2021. doi: 10.3389/fninf.2021.596443

[11] S. Uda, "Application of information theory in systems biology," *Biophys. Rev.*, vol. 12, pp. 377–384, 2020. doi: 10.1007/s12551-020-00665-w

[12] N. Amer and S. Belhaouari, "EEG Signal Processing for Medical Diagnosis, Healthcare, and Monitoring," *IEEE Access*, vol. 11, pp. 143116–143142, 2023. doi: 10.1109/access.2023.3341419

[13] T. Wang, W. Jia, F. Li, X. Liu, X. Zhang, and F. Hu, "Multi-dimensional feature extraction of EEG signal and its application in stroke classification," *Sci. Rep.*, vol. 15, 2025. doi: 10.1038/s41598-025-04756-0

[14] F. Von Wegner, M. Wiemers, G. Hermann, I. Tödt, E. Tagliazucchi, and H. Laufs, "Complexity Measures for EEG Microstate Sequences: Concepts and Algorithms," *Brain Topogr.*, vol. 37, pp. 296–311, 2023. doi: 10.1007/s10548-023-01006-2

[15] Z. Zhang, "Research on Mutual Information Analysis in EEG Based on Different Entropy Measures," in Proc. IEEE 5th Int. Conf. Information Systems and Computer Aided Education (ICISCAE), 2022, pp. 548–553. doi: 10.1109/iciscae55891.2022.9927610.

[16] Asgharzadeh-Bonab, M. Amirani, and A. Mehri, "Spectral entropy and deep convolutional neural network for ECG beat classification," *Biocybern. Biomed. Eng.*, vol. 40, pp. 691–700, 2020. doi: 10.1016/j.bbe.2020.02.004

[17] T. Li and M. Zhou, "ECG Classification Using Wavelet Packet Entropy and Random Forests," *Entropy*, vol. 18, no. 8, p. 285, 2016. doi: 10.3390/e18080285

[18] H. Wu, "Multiscale entropy with electrocardiograph, electromyography, electroencephalography, and photoplethysmography signals," *Biomed. Signal Process. Control*, vol. 93, p. 106124, 2024. doi: 10.1016/j.bspc.2024.106124

[19] M. Coq, A. Rey, O. Rosso, R. Armentano, and W. Legnani, "Detection of cardiac arrhythmia patterns in ECG through H × C plane," *Chaos*, vol. 32, no. 12, p. 123118, 2022. doi: 10.1063/5.0118717

[20] S. Aulia and S. Hadiyoso, "Abnormal ECG Classification using Empirical Mode Decomposition and Entropy," *Jurnal Rekayasa Elektrika*, vol. 17, no. 3, 2021. doi: 10.17529/jre.v17i3.22070

[21] L. Xie, Z. Li, Y. Zhou, Y. He, and J. Zhu, "Computational Diagnostic Techniques for Electrocardiogram Signal Analysis," *Sensors*, vol. 20, no. 21, p. 6318, 2020. doi: 10.3390/s20216318

[22] E. Ramírez, S. Ruipérez-Campillo, R. Casado-Arroyo, J. Merino, J. Vogt, F. Castells, and J. Millet, "The art of selecting the ECG input in neural networks to classify heart diseases," *Front. Physiol.*, vol. 15, 2024. doi: 10.3389/fphys.2024.1452829

[23] J. Sahores, V. Bank, M. Meijome, I. Ziccardi, and W. Legnani, "Study of Information Flow Among ECG Leads," in IEEE Biennial Congress of Argentina (ARGENCON), 2024, pp. 1–8. doi: 10.1109/argencon62399.2024.10735814

[24] B. Andayeshgar, F. Abdali-Mohammadi, M. Sepahvand, A. Almasi, and N. Salari, "Arrhythmia detection by the graph convolution network and a proposed structure for communication between cardiac leads," *BMC Med. Res. Methodol.*, vol. 24, 2024. doi: 10.1186/s12874-024-02223-4

[25] H. Zhang, Z. Bian, X. Guo, X. Li, S. Yin, and D. Cui, "Measuring phase-amplitude coupling using dispersion fuzzy mutual information," *Comput. Methods Programs Biomed.*, vol. 272, p. 109075, 2025. doi: 10.1016/j.cmpb.2025.109075

[26] L. A. Luengas, "Modelo de alineación estática para prótesis transtibiales", PhD Thesis, Facultad de Ingeniería, Pontificia Universidad Javeriana, 2016. https://repository.javeriana.edu.co/items/4297eb56-d54f-401c-bb6c-8cb1b7c029a1

[27] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation," 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 2018, pp. 443–444. doi: 10.1109/ICHI.2018.00092.

[28] Pyle, D. (1999). Data preparation for data mining. In Morgan Kaufmann Publishers, Inc. Morgan Kaufmann Publishers, Inc. https://doi.org/10.1080/713827180

FUNDING

Product derived from the research project entitled "Development of an application for the early detection of atrial fibrillation based on digital signal processing techniques applied to photoplethysmography (PPG) data". Project funded by the Sistema Universitario Estatal (SUE), under project code PPEE-CV SUE DC 02 2023.

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

All authors have read and approved the final version of the manuscript.