

Advancing ECG Signal Analysis with a Novel Geometric Feature Space for Differentiating Atrial Fibrillation

H. GARCÍA BLESA^{1*}, J. VOROBIOFF² and W. LEGNANI³

Received on May 13, 2025 / Accepted on April 13, 2026

ABSTRACT. Electrocardiogram (ECG) signals are inherently challenging to classify due to their complex and variable nature. This study introduces a novel approach for classifying normal and atrial ECG signals by utilizing a geometric feature space (GFS) generated from three specific geometric characteristics: amplitude, Zenith angle, and shape factor. By employing a sweeping technique to extract and compile unique information, the proposed algorithm builds this data-set, which is subsequently processed by standard machine learning models. To frame the research contribution, the GFS is quantitatively benchmarked against state-of-the-art techniques, specifically addressing the performance limitations of non-linear dynamics and deep learning models (CNN/LSTM). Our methodology demonstrates substantial improvements in classification performance, achieving an average F1-score of 0.98. These results underscore the effectiveness and robustness of our approach in distinguishing between normal and atrial ECG signals.

Keywords: amplitude, zenith angle, shape factor, differentiation of signals, electrocardiogram signals, training and testing.

1 INTRODUCTION

Signal processing is central to almost any modern ECG analysis system and has proven to be essential to improve the diagnosis of a wide range of cardiac pathologies. This process is used to solve multiple problems in ECG analysis, such as data compression, beat detection and classification, noise removal, signal separation, and feature extraction, as explained in [6].

*Corresponding author: H. García Blesa – E-mail: hgarciablesa@frba.utn.edu.ar

¹Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Signal and Image Processing Center, Avenida Medrano 951 (C1179AAQ), Ciudad Autónoma de Buenos Aires, Argentina – E-mail: hgarciablesa@frba.utn.edu.ar
<https://orcid.org/0000-0003-0491-2018>

²Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Signal and Image Processing Center, Avenida Medrano 951 (C1179AAQ), Ciudad Autónoma de Buenos Aires, Argentina – E-mail: jvorobioff@frba.utn.edu.ar
<https://orcid.org/0000-0001-9346-061X>

³Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Signal and Image Processing Center, Avenida Medrano 951 (C1179AAQ), Ciudad Autónoma de Buenos Aires, Argentina – E-mail: walter@frba.utn.edu.ar
<https://orcid.org/0000-0002-6949-0728>

A cardiac disorder that is particularly difficult to diagnose is atrial fibrillation (AF), which is characterized by an irregular and often abnormally fast heart rhythm as pointed out in [15]. This disorder affects about 10% of persons over 75 years old as noted in [1]. In this condition, characterized by sporadic electrical activity in the atria which depolarize between 300 and 600 beats per minute. No effective contraction of the atria results, rather causing a muscle tissue ripple called fibrillation as defined by [19].

The last few years have marked a renaissance in the use of machine learning (ML) techniques for healthcare innovation [14]. This resurgence comes against the backdrop of the emerging era of electronic health records, with structured collections of digitized medical data, enhanced by the power of advanced analysis tools. Along with these advancements, several technological improvements in hardware, such as cloud computing, increased computational power, and the use of GPUs, have played a key role as is suggested by [13]. These developments, combined with modern software efforts, have propelled the progress of ML and deep learning technologies, which rely on data to automatically identify patterns for specific tasks. These newer technologies have achieved great success in other areas like speech recognition and image analysis, and are extensively developed in the literature ([5], [4] and [10]). In clinical electrophysiology, the storage and organization of large digitized datasets, along with the availability of high-performance computing platforms, have enabled the integration and adoption of machine learning technologies to optimize the extraction of information from extensive ECG datasets as highlighted by [11].

Under normal conditions, the heart contracts and relaxes rhythmically to pump blood; however, in the presence of AF, the atria contract randomly and rapidly, reducing the heart's efficiency and making it difficult for the atria to relax properly between contractions. This complexity is further heightened by the paroxysmal and asymptomatic nature of some AF cases, which complicates its detection in routine examinations, as highlighted in [17] and [2]. This condition can progress into a serious clinical scenario that demands intensive monitoring, as noted by [19].

The proposed algorithm for the classification of ECG signals with AF employs the sweep technique, extracting geometric features of the signal at each step. Using this information, it constructs a feature space that is subsequently processed with machine learning algorithms to differentiate signals and detect pathology, as developed in [16] and [18]. The algorithms applied in this research work include Support Vector Classifier, Gaussian Process Classifier, Decision Tree Classifier, Random Forest Classifier, Multi-Layer Perceptron, Adaptive Boosting Classifier, Gaussian Naive Bayes, Quadratic Discriminant Analysis, Logistic Regression, and K-Neighbors Classifier. To evaluate classification performance, parameters usually used were standard F1-score, balanced accuracy, sensitivity, specificity, positive predictive value, and negative predictive value.

The main contributions of this paper are centered on the specific definition and application of the GFS to ECG signal analysis for atrial fibrillation detection. While the underlying extraction mechanism, sweeping technique, and general processing pipeline follow the framework established in [8], the transformation of ECG signals into a GFS—through the combination of amplitude, zenith angle, and shape factor—and the subsequent demonstration of its superior perfor-

mance over deep learning models and non-linear dynamics constitute the primary innovation of this research. This approach seeks to address the performance limitations observed in traditional statistical or temporal aggregation techniques.

2 METHOD

The general scheme used by this method is to process a database of discrete signals using the [8] algorithm to generate the feature space. Subsequently, the feature space becomes the input data for the machine learning algorithms. Figure 1 shows this simple structure.

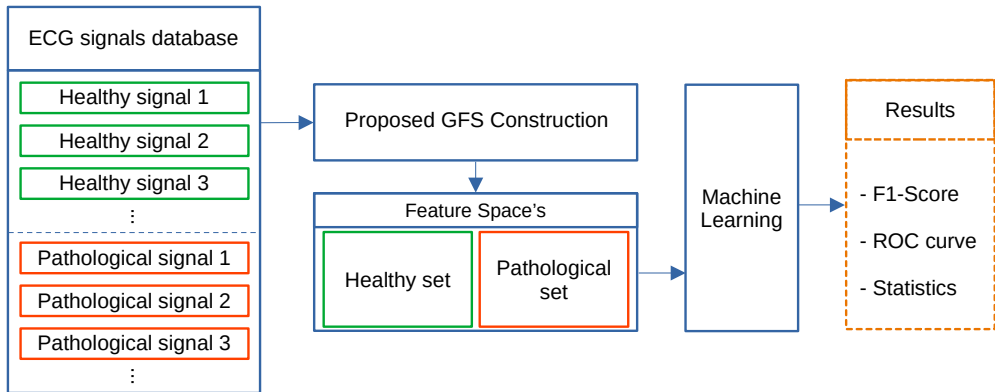


Figure 1: General scheme of the method.

Below, we provide a brief review of the key elements presented and the feature space used.

The amplitude of the signal S , denoted as $\min(S)$ for the minimum value and $\max(S)$ for the maximum value, is calculated as follows:

$$A(S) = \max(S) - \min(S). \quad (2.1)$$

The next parameter, called the Zenith angle, is an application of the average of the values in S . For this, the mean value p_S of the values in S is used to observe how it lies within the bounded space between $\min(S)$ and $\max(S)$. The following set in $\mathbb{R}^2$ is considered: $\mathcal{S}_S = [0, 1] \times [\min(S), \max(S)]$. The angle φ_S determined by the segments $\overline{PM}$ and $\overline{PN}$, where $P = (0.5, p_S)$, $M = (1, \max(S))$, and $N = (1, p_S)$, is defined. This definition is illustrated in Figure 2a. Then, in order to compare φ_S values from different signals, the set $\mathcal{S}_S$ is normalized as $\mathcal{S} = [0, 1] \times [0, 1]$, and the average p_S is normalized.

$$p = \frac{p_S - \min(S)}{\max(S) - \min(S)}. \quad (2.2)$$

Finally, we calculate:

$$\varphi = \arctan\left(\frac{1-p}{0.5}\right) = \arctan(2-2p). \quad (2.3)$$

Figure 2b shows the normalized value of φ .

Since the average cannot be less than the minimum nor greater than the maximum, the values will range between 0 and 1, determining the bounds of φ . When $p = 0$, it is found that $\varphi = \arctan(2) \leq 64^\circ \cong 1.1071$ rad, and when $p = 1$, it is determined that $\varphi = \arctan(0) = 0^\circ$. Thus, $\varphi \in [0, 1.1071)$, or equivalently, $\varphi \in [0, 63.43^\circ)$ when expressed in sexagesimal degrees. It is evident that the Zenith angle is highly sensitive to variations in the values of the signal S . When a value in S is significantly close to the maximum, the average tends to disregard the remaining values and shifts toward the upper extreme. Conversely, if a value is very close to the minimum, the angle shifts toward the opposite extreme. In contrast, when the values exhibit minimal variation, the average remains near the center.

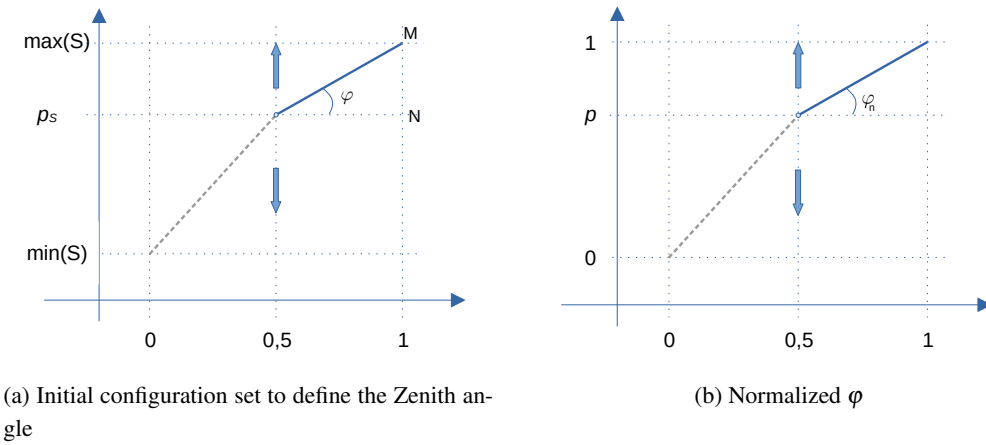


Figure 2: Representation of the Zenith angle.

It is proposed that the signal be described in terms of peaks and valleys. For the formulation of the following parameter, an investigation was carried out on how a measurement could be extracted from the shape of the signal, with the idea of condensing it into a single value. To achieve this, the following calculation is proposed. Let the signal $S_i = \{a_1, a_2, \dots, a_n\}$ be considered. First, the data is arranged in increasing order, producing the sequence $B = \{b_1, b_2, \dots, b_n\} \subseteq S_i$, where $b_i \leq b_{i+1}$ for $i = 1, 2, \dots, n - 1$. Repeated elements in B are then removed, resulting in the sequence $C = \{c_1, c_2, \dots, c_m\} \subseteq S_i$, where $m \leq n$ and $c_i < c_{i+1}$ for $i = 1, 2, \dots, m - 1$. Finally, the vector $O = (o_1, o_2, \dots, o_n)$ is defined such that $o_i = j$ if $a_i = c_j$, for $i = 1, 2, \dots, n$.

When generating the vector O , it should be noted that if repeated values are present in the sequence under analysis, their respective order numbers will also be repeated. This particularity implies that the components of the ordered vector are bounded between 1 and the length of the original sequence. The shape of the sequence in terms of peaks and valleys from positive integers is represented by the vector O . In order to quantify and synthesize this vector, a number will be constructed using the components of O as exponents of a prime number base. Explicitly, let

$\mathcal{P}_n = \{q_1, q_2, \dots, q_n\}$ be defined as the set of the first n prime numbers, and the shape parameter is defined as follows:

$$\mathcal{F}(S) = \prod_{i=1}^n q_i^{o_i}. \quad (2.4)$$

The Fundamental Theorem of Arithmetic guarantees the deconstruction of the shape parameter to recover the valley and peak structure of the sequence. An example is presented in the following to show the previously described procedure. Next, a segment of the signal is defined where the embedding technique is applied. Considering a discrete signal S of length n , and given an embedding dimension $d < n$, the following segments of length d derived from S are defined:

$$S_i = \{a_i, a_{i+1}, \dots, a_{i+d-1}\}, \quad (2.5)$$

for $i = 1, 2, \dots, n - d + 1$. Note that the segment S_i is itself a signal, allowing all previously described techniques to be applied to it. Each segment S_i , as defined by expression (2.5), can be associated with a term determined by

$$S_i \mapsto (A(S_i), \varphi(S_i), \mathcal{F}(S_i)). \quad (2.6)$$

In order to simplify notation, we denote $A_i = A(S_i)$, $\varphi_i = \varphi(S_i)$, and $\mathcal{F}_i = \mathcal{F}(S_i)$. Replacing in the expression (2.6), we define the vector

$$v_i = (A_i, \varphi_i, \mathcal{F}_i). \quad (2.7)$$

In order to obtain a fingerprint, or to obtain features that highlight particular properties, the eventual repetition of vectors v_i reveals a redundancy of information. To get a reduced set that shows unique characteristics it is necessary to remove the redundant information. The first two components of the vector v_i –amplitude and Zenith angle– are real values, so the probability of their occurrence is considered very low. To increase this probability, both parameters are discretized, and the calculated values are assigned to a discrete range of arbitrary size. The total amplitude of the signal, $\max(A)$, is obtained and divided by an arbitrary value $t \in \mathbb{N}$. A set $\mathcal{A}$ with t elements is thus generated, defined as:

$$\mathcal{A} = \{A'_j : A'_j = \max(A)/j \text{ for } j = 1, 2, \dots, t\}. \quad (2.8)$$

Then, A_i is bounded by two values of $\mathcal{A}$. In other words, there exists j_i between 1 and $t - 1$ such that $A'_{j_i} \leq A_i \leq A'_{j_i+1}$. Finally, after calculating A_i , the discrete amplitude A'_{j_i+1} is assigned to the first component of the feature vector. It is clear that the coincidence of the values of the discretized amplitude depends on the value of t .

Analogously, a discrete set of angle values is defined. As indicated in 2.3, φ has a maximum value. Given the same arbitrary value t , the set is constructed as

$$\mathcal{Z} = \{\varphi'_j : \varphi'_j = \max(\varphi)/j, j = 1, 2, \dots, t\}. \quad (2.9)$$

Each value φ_i is then bounded by two consecutive elements of $\mathcal{Z}$. That is, there exists an index $j_i \in \{1, \dots, t-1\}$ such that

$$\varphi'_{j_i} \leq \varphi_i \leq \varphi'_{j_i+1}.$$

The feature vector with the parameters amplitude and Zenith angle discretized is re-defined as $v_i = (A'_i, \varphi'_i, \mathcal{F}_i)$ where A'_i is the discrete amplitude from A_i as well as φ'_i corresponds to the discrete value of φ_i . The set of all these vectors is $V = \{v_i : i = 1, 2, \dots, n-d+1\}$. It is noted that discretization depends on the value of t . For $t = 1$, v_i is defined in (2.7). For the purpose of filtering out redundant information, the set V is reduced as much as possible by removing repeated elements. Thus, the set $V' = \{v'_1, v'_2, \dots, v'_k\}$ is considered, where $v_i \in V$ for $i = 1, 2, \dots, k$, and for all i, j between 1 and k , such that $i \neq j$, it holds that $v_i \neq v_j$.

This set, consisting of unique vectors, has cardinality $\#V' \leq \#V$. It is relevant to know the frequency in V of the vectors of V' . For $i = 1, 2, \dots, k$, the frequency $F_i = \#\{v \in V : v = v_i\}$ is defined. Finally, the set is defined as follows:

$$W = \{(v'_i, F_i) : i = 1, 2, \dots, k\}, \quad (2.10)$$

containing the vectors of V' and their respective frequencies in V . Thus, W set is the GFS. If $\mathcal{S} = \{S_1, S_2, \dots, S_t\}$ is a set of signals, set W is associated with $\mathcal{S}$ as its GFS.

The experiment focused on the use of the GFS as a tool for classifying discrete signals. The algorithm generates the GFS for each group of signals and formats it into a structured and labeled table suitable for processing by a standard classifier. A short example of GFS file format is shown in Table 1. The algorithm builds one single file by adding the feature vectors of both groups of signals including its corresponding labels.

The signal database consists of 416 signals, including 282 normal and 134 atrial fibrillation cases. The signals were labeled as 'NOR' (normal) and 'ATR' (atrial fibrillation). These data were obtained from PhysioNet. In this study, we used the same dataset as in [19], namely the MIT-BIH Arrhythmia Database available at <http://www.physionet.org>. It includes recordings from 23 patients with normal sinus rhythm and 6 patients with atrial fibrillation, covering both healthy individuals and subjects with arrhythmias. A total of 3600 non-overlapping samples were extracted from 1000 randomly selected ECG segments. The signals were recorded at a sampling rate of 360Hz with a gain of 200 adu/mV in the primary ECG lead, representing 45 subjects: 19 women (aged 23 to 89) and 26 men (aged 32 to 89). Figure 3 shows a sample of a normal ECG signal.

We performed the differentiation calculations for two groups of ECG signals: normal signals (NOR) and signals with atrial fibrillation (ATR) using our algorithm. The GFS sets was used and post-processing was performed using the aforementioned classifiers. The models chosen to carry out this process are widely disseminated in the literature, see [12] for more details. *Support Vector Classifier (SVC)*, a classifier that finds the best hyperplane to separate classes in the data. The configuration used for the experiment was kernel = "linear", $C = 0.025$. *Support Vector Classifier (SVC)*: Similar to the previous one but uses a nonlinear kernel to capture more complex relation-

Table 1: A short example of the GFS's generated by de model. Every row is a feature vector plus frequency. First column indicate the GFS. The second column is the label of the group, the third is the amplitude obtained from a segment, the fourth is the Zenith angle, the fifth is the shape factor and the last column is the frequency of this 3-tuple found.

<i>Feature-space</i>	Label	Amplitude	Zenith angle	Shape factor	Freq.
<i>ATR group</i>	ATR	8.041	54.554	9147600	3
	ATR	8.041	49.479	23478840	2
	ATR	16.082	57.091	831600	2
	⋮	⋮	⋮	⋮	⋮
	ATR	16.082	59.629	29348550	1
<i>NOR group</i>	NOR	7.612	54.554	970200	3
	NOR	7.612	58.360	8385300	1
	NOR	15.224	60.898	903935340	1
	⋮	⋮	⋮	⋮	⋮
	NOR	7.612	50.748	242550	4

ships. The configuration used for the experiment was $\gamma = 2, C = 1$. *Gaussian Process Classifier (GPC)*: A probabilistic classifier that models distributions over functions using Gaussian processes. The configuration used for the experiment was $1.0 \times \text{RBF}(1.0)$. *Decision Tree Classifier (DTC)*: A classifier that uses a decision tree to split the data based on features. The configuration used for the experiment was $\text{max_depth} = 5$. *Random Forest Classifier (RFC)*: An ensemble of decision trees that improves accuracy by majority voting from multiple trees. The configuration used for the experiment was $\text{max_depth} = 5, \text{n_estimators} = 10, \text{max_features} = 1$. *Multi-Layer Perceptron Classifier (MLP)*: A neural network with multiple layers that adjusts weights using gradient descent. The configuration used for the experiment was $\alpha = 1, \text{max_iter} = 1000$. *AdaBoost Classifier (AdaBoost)*: An ensemble algorithm that combines several weak classifiers into a strong model by adjusting the weights of the examples. *Gaussian Naive Bayes (GNB)*: A probabilistic classifier that assumes conditional independence between features and models the distribution with a Gaussian. *Logistic Regression (LR)*: A regression model used for binary classification that estimates the probability of class membership. *K-Neighbors Classifier (KNN)*: A classifier that assigns the majority class of the k nearest neighbors in the feature space. The configuration used for the experiment was $k = 3$.

We calculated the following standard parameters for each model: F1-score, balanced accuracy, sensitivity, specificity, PP values, NP values, and the processing time. If the F1-score should prove to be high and representative of good classification, all the other parameters would show positive results. This would lead to high balanced accuracy, which would mean sensitivity and specificity were also high. Sensitivity, or recall, would be expected to be high, as a high F1-score would indicate that the model was effective at identifying positive instances, meaning that true positives made up a large proportion of the total actual positives. Specificity would also be high;

normally, high recall would refer to being accurate about negative instances, assuming the model were properly calibrated. In this case, both the PP value and precision should prove to be high because the F1-score, being the harmonic mean of precision and recall, should prove to show high precision for the elevated F1-score. The NP value should prove to remain high because effective classification of the positive class usually leads to good performance in classifying the negative class, assuming the model is balanced, as discussed by [3], [9], and [7]. A more thorough assessment of the machine learning methods was ensured by using the K-fold cross-validation function to reduce biases in the analysis as has been observed in [20]. To ensure reproducibility, the configuration specifically included splitting the dataset into 5-folds ([20]), shuffle the data before to splitting, and establishing a fixed random seed. The parameters computed across all folds were averaged to produce the final findings.

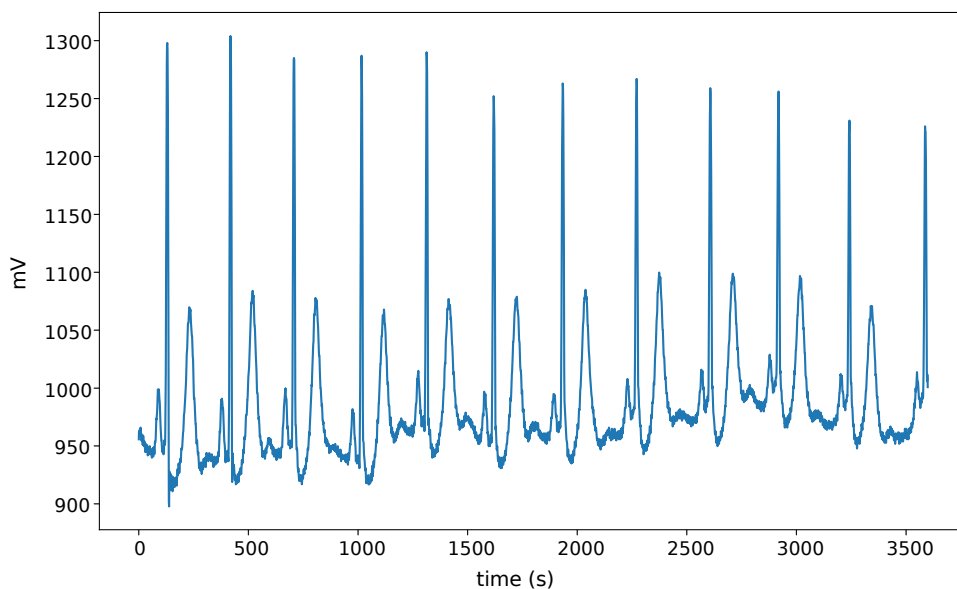


Figure 3: ECG signal from a normal case. Resolution: 3600 points in 10 seconds (≈ 2.78 ms/point).

The initial parameters for our algorithm were defined arbitrarily in two sets: the set D corresponding to embedding dimension values, $D = \{3, 4, 5\}$, and the set T of discretization factor values, $T = \{50, 75, 100\}$. We denote a pair of input parameters as $[d, t]$. A total of 9 cases were calculated by combining elements from both sets D and T obtaining 9 pairs of input parameters $[3, 50], [3, 75], [3, 100], [4, 50], [4, 75], [4, 100], [5, 50], [5, 75], [5, 100]$, generating a respective feature space for each case. Then, the algorithms from the list were applied to each generated feature space, resulting in a total of 99 results.

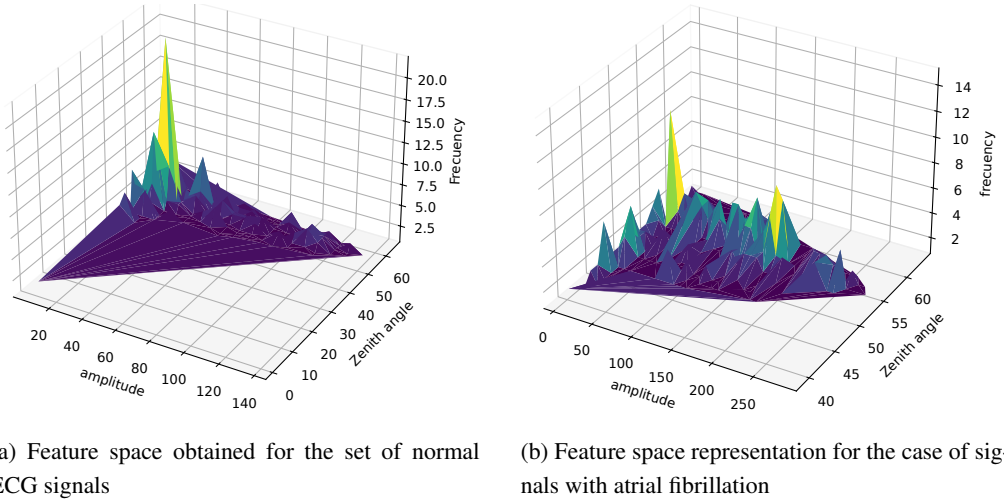


Figure 4: This two graphs shows the internal structure of the feature spaces obtained by our model for the case $[5, 50]$.

3 RESULTS

Of the total results from the classification process using the aforementioned ML models, 61 of them showed negative overfitting, thereby ruling out the results of the remaining 27 with positive overfitting. Table 2 shows the results for the models that obtained F1-score values above 90%, among which there are only two: AdaBoost and Decision Tree. The best results were obtained by the AdaBoost model, with the four best F1-scores ranging from 0.977 to 0.959 in the cases $[d, t] = [5, 50]$, $[d, t] = [5, 75]$, $[d, t] = [5, 100]$, and $[d, t] = [4, 50]$. In particular, the results for this model are shown, as the remaining F1-scores recorded values between 85% and 89%. The second-best performing model was Decision Tree, with F1-scores of $[d, t] = [5, 50]$, $[d, t] = [5, 75]$, and $[d, t] = [5, 100]$.

The Balanced Accuracy, Sensitivity, and Specificity metrics remain within ranges close to the F1-score values in most configurations. In some cases for the Decision Tree model, the Specificity, PPV, and NPV values reach 1.

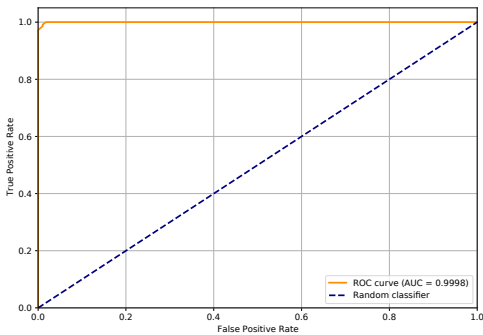
The difference between F1-tr and F1-test remains small across all evaluated configurations, with values ranging from 0 to 0.04638. In the column corresponding to the overfitting indicator, all configurations are recorded as 'NO'.

4 CONCLUSIONS

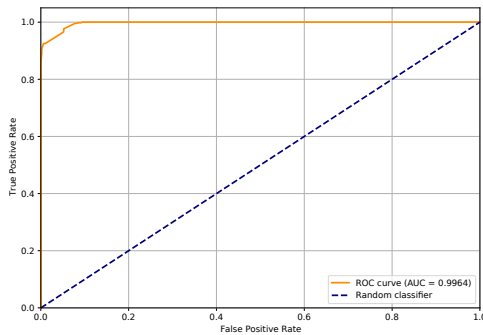
In this experiment, the GFS was used for comparisons of normal and atrial ECG signals. The main purpose of using this space was to test whether this tool would improve the discrimination between the two types of signals. The results indicate that the differentiation is successful when

Table 2: Differentiation by average results using classifiers. Only results with F1-score mean greater than 0.9 are shown.

D	K	Modelo	F1 Score	Bal. Acc.	Sens.	Spec.	PPV	NPV	F1-tr	F1-test	dif-F1	Overfitting
5	50	AdaBoost	0.9774	0.9775	0.9818	0.9733	0.9732	0.9733	0.9853	0.9774	0.0079	NO
5	75	AdaBoost	0.9709	0.9711	0.9789	0.9633	0.9622	0.9633	0.9750	0.9709	0.0041	NO
5	100	AdaBoost	0.9617	0.9618	0.9683	0.9553	0.9544	0.9553	0.9674	0.9617	0.0057	NO
4	50	AdaBoost	0.9600	0.9602	0.9585	0.9619	0.9596	0.9619	0.9735	0.9600	0.0135	NO
5	50	Decision Tree 5	0.9585	0.9589	0.9973	0.9206	0.9250	0.9206	0.9598	0.9585	0.0012	NO
4	75	AdaBoost	0.9516	0.9524	0.9669	0.9378	0.9338	0.9378	0.9589	0.9516	0.0073	NO
3	50	AdaBoost	0.9441	0.9420	0.9235	0.9604	0.9555	0.9604	0.9905	0.9441	0.0464	NO
5	75	Decision Tree 5	0.9384	0.9373	0.8746	1.0000	1.0000	1.0000	0.9384	0.9384	0.0000	NO
5	100	Decision Tree 5	0.9351	0.9343	0.8686	1.0000	1.0000	1.0000	0.9351	0.9351	0.0000	NO
4	100	AdaBoost	0.9268	0.9273	0.9502	0.9045	0.9035	0.9045	0.9438	0.9268	0.0170	NO
4	50	Decision Tree 5	0.9117	0.9147	0.9892	0.8402	0.8541	0.8402	0.9182	0.9117	0.0064	NO



(a) ROC curve corresponding to F1-score average 0.9774, 4th-fold



(b) ROC curve corresponding to F1-score average 0.9709, 4th-fold

Figure 5: Two of the five ROC curves calculated for the top average F1-score obtained by the cases [5, 50] (left) and [5, 100] (right).

employing the GFS. This improvement highlights the effectiveness of the GFS in capturing and distinguishing the unique characteristics of normal and atrial ECG signals. Additionally, the classification of the groups is consistent, as evidenced by the high values of other parameters such as balanced accuracy, sensitivity, specificity, positive predictive value, and negative predictive value.

These consistently high values across different metrics further validate the robustness and reliability of the classification performance. This improvement will clearly show how the GFS is effective in capturing and distinguishing between unique characteristics of normal and atrial ECG signals.

An additional advantage of the employed feature space is its ability to be visualized (as shown in Figure 4). This visualization reveals a difference between the two spaces, characterized by the peaks present in the ATR space. This distinction is another feature of the present model that

underscores the effectiveness of the feature space in differentiating between normal and atrial ECG signals.

The experiment presented is comparable with the one conducted by Ziccardi et al. [19]. The referenced study performed a similar analysis by generating multiple feature spaces using various tools such as Rényi entropy, statistical complexity, Fisher information, Lyapunov exponent, and fractal dimension. These feature spaces were then processed by similar models used in our work, achieving a maximum F1-score of 0.79. Additionally, in [20], the authors employed an ML technique to classify ECG signals with AF. In this case, the authors employed a deep convolutional neural network (CNN) with averaging-based feature aggregation across time and convolutional layers for feature extraction, along with long short-term memory (LSTM) layers for temporal aggregation of features. They measured the difference between the two groups using the same statistical parameters, obtaining a maximum F1-score of 0.82. In contrast, our approach achieved a significantly higher maximum leading to an average F1-score of approximately 0.98 (0.9774 over 5-fold cross-validation). This substantial improvement underscores the efficacy and robustness of our method, highlighting the advancements and enhancements introduced in our study.

Acknowledgments

To the Secretariat of Science and Technology and the Signal and Image Processing Center of the National Technological University for their support and collaboration in this research.

Data availability

The data that support the findings of this study is MIT-BIH Arrhythmia Database, available at <http://www.physionet.org>.

Associate editor: Paulo Pagliosa

REFERENCES

- [1] R. Abumuaileq. Atrial fibrillation in the old/very old: prevalence and burden, predisposing factors and complications. *Curr Res Med*, **17** (2019), 1–10.
- [2] O. Berenfeld, V.D. Corino, A. Loewe, J.P. Martínez & J.F. Rodríguez Matas. Atrial Fibrillation: Technology for Diagnosis, Monitoring, and Treatment (2022).
- [3] C.M. Bishop. Pattern recognition and machine learning. *Springer google schola*, **2** (2006), 1122–1128.
- [4] A. Bose & B. Tripathy. Deep learning for audio signal classification. *Deep learning research and applications*, (2020), 105–136.
- [5] F. Camastra & A. Vinciarelli. “Machine learning for audio, image and video analysis: theory and applications”. Springer (2015).

- [6] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann & J.M. Roig. Principal component analysis in ECG signal processing. *EURASIP Journal on Advances in Signal Processing*, **2007** (2007), 1–21.
- [7] D. Chicco & G. Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, **21** (2020), 1–13.
- [8] M. García Blesa, J. Vorobioff & W. Legnani. New Methodology for Differentiating Noisy Signals of Different Colours Using Geometric Features. In “2024 IEEE Biennial Congress of Argentina (ARGENCON)” (2024), p. 1–5. doi:10.1109/ARGENCON62399.2024.10735928.
- [9] A. Géron. “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow”. O’Reilly Media, Inc.” (2022).
- [10] U. Kamath, J. Liu & J. Whitaker. “Deep learning for NLP and speech recognition”, volume 84. Springer (2019).
- [11] A. Mincholé, J. Camps, A. Lyon & B. Rodríguez. Machine learning in the electrocardiogram. *Journal of Electrocardiology*, **57** (2019), S61–S64. doi:<https://doi.org/10.1016/j.jelectrocard.2019.08.008>. URL <https://www.sciencedirect.com/science/article/pii/S0022073619304571>.
- [12] K.P. Murphy. “Machine Learning: A Probabilistic Perspective”. MIT Press (2012).
- [13] M. Pandey, M. Fernandez, F. Gentile, O. Isayev, A. Tropsha, A.C. Stern & A. Cherkasov. The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence*, **4**(3) (2022), 211–221.
- [14] A. Rajkomar, E. Oren, K. Chen, A.M. Dai, N. Hajaj, M. Hardt, P.J. Liu, X. Liu, J. Marcus, M. Sun *et al.* Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, **1**(1) (2018), 18. doi:10.1038/s41746-018-0029-1. URL <https://doi.org/10.1038/s41746-018-0029-1>.
- [15] F. Richard Espiga, M. Almendro Delia, F. Caballero Martínez, D. Monge Martín, F. Neria Serrano & R. Quirós López. Delayed diagnosis and missed opportunities in the early detection of atrial fibrillation: a cross-sectional study. *Revista Clínica Española (English Edition)*, **224**(9) (2024), 560–568. doi: <https://doi.org/10.1016/j.rceng.2024.08.002>. URL <https://www.sciencedirect.com/science/article/pii/S2254887424001115>.
- [16] I.H. Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, **2**(3) (2021), 160.
- [17] Y.C. Wang, X. Xu, A. Hajra, S. Apple, A. Kharawala, G. Duarte, W. Liaqat, Y. Fu, W. Li, Y. Chen *et al.* Current advancement in diagnosing atrial fibrillation by utilizing wearable devices and artificial intelligence: A review study. *Diagnostics*, **12**(3) (2022), 689.
- [18] Ö. Yıldırım, P. Plawiak, R.S. Tan & U.R. Acharya. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in biology and medicine*, **102** (2018), 411–420.
- [19] I. Ziccardi, A. Rey & W. Legnani. ECG signals classification using overlapping variables to detect atrial fibrillation. *Trends in Computational and Applied Mathematics*, **23** (2022), 569–581.

- [20] M. Zihlmann, D. Perekrestenko & M. Tschannen. Convolutional recurrent neural networks for electrocardiogram classification. In “2017 Computing in Cardiology (CinC)” (2017), p. 1–4. doi:10.22489/CinC.2017.070-060.

How to cite

H. García Blesa, J. Vorobioff & W. Legnani. Advancing ECG Signal Analysis with a Novel Geometric Feature Space for Differentiating Atrial Fibrillation. *Trends in Computational and Applied Mathematics*, **27**(2026), e01870. doi: 10.5540/tcam.2026.027.e01870.

