*Neutrosophic Sets and Systems, {Special Issue: Neutrosophic and Plithogenic Approaches in Data Science and Multivariate Analysis: Contributions from the IX Ibero-American Biometry Meeting}, Vol. 89, 2025*

*University of New Mexico*

# Predictive Modeling of Atrial Fibrillation Using a Hybrid Multinomial-Neutrosophic Approach for Biomarker Identification

**Lorenzo Cevallos-Torres[1,2,*], Luis Pilacuan-Bonete[3], Rosangela Caicedo-Quiroz[1], Franklin Parrales-Bravo[1,2], Eduardo Rubio-Bonito[2]**

[1]Centro de Estudios en Tecnologías Aplicadas, Universidad Bolivariana del Ecuador, Durán, Ecuador, ljcevallost@ube.edu.ec, frparralesb@ube.edu.ec, rcaicedoq@ube.edu.ec

[2]Grupo de Investigación en Inteligencia Artificial, Universidad de Guayaquil, lorenzo.cevallost@ug.edu.ec, franklin.parralesb@ug.edu.ec

[3]Facultad de Ingeniería Industrial, Universidad de Guayaquil, Ecuador. luis.pilacuanb@ug.edu.ec

**Abstract:** Atrial fibrillation, characterized by chaotic rhythms and electrical complexity, presents a diagnostic challenge that requires innovative approaches to uncover its underlying biomarkers. This study proposes a hybrid predictive model based on multinomial logistic regression and neutrosophic logic, aiming to identify clinically significant patterns associated with this condition. Using the Knowledge Discovery in Databases (KDD) methodology, large volumes of cardiovascular data are analyzed to distinguish meaningful signals from background noise, revealing hidden connections and validating medical hypotheses. The implementation of the model through a digital prototype reflects a convergence of advanced statistics, artificial intelligence, and cardiovascular medicine, promoting a multidisciplinary approach. The findings of this work not only enhance diagnostic accuracy but also open new avenues for personalized treatment, emphasizing the value of scientific integration in modern medical research.

**Keywords:** Atrial fibrillation; Electrocardiogram; Predictive models; Biomarkers; Artificial intelligence; Clinical data.

## 1. Introduction.

Atrial fibrillation (AF) manifests as a complex cardiac symphony involving considerable risks such as stroke and heart failure, while casting a shadow on quality of life [1]. Although its diagnosis and treatment manifest as colossal challenges due to the complex structure of constituent factors, Atrial Fibrillation emerges as the central focus of this study. Through a bold multidisciplinary methodology, integrating comprehensive literature reviews, rigorous clinical analyses, and the implementation of innovative predictive models, this study delves into the depths of the Faculty of Care [2]. The interrelationship between the fields of cardiology, medical informatics, and genetics manifests itself as the guiding beacon, illuminating the winding paths that envelop this cardiac condition. In its glow, one senses the promise of unexplored routes to personalized interventions and optimized outcomes for individuals immersed in the intricate web of this condition.

## 2. Materiales y Métodos.

### 2.1. Data collection and literature review

A comprehensive scientific evaluation was performed to collect the most pertinent information in the expansive domain of cardiology. This encompassed the examination of prominent journals including Circulation, European Heart Journal, and Journal of the American College of Cardiology [3]. Furthermore, pertinent fields such as medical informatics and genetics were examined, utilizing resources such as Procedia Computer Science and Frontiers in Genetics. Key variables related to atrial fibrillation (AF) were identified through this technique [4]. The selected variables were meticulously arranged, yielding a solid and valuable dataset that will underpin future scientific research.

### 2.2. Acquisition and structuring of datasets.

The Kaggle digital platform was utilized to get two critical datasets: one comprising general patient information and the other consisting of twelve-lead electrocardiogram (ECG) recordings [5]. The datasets were amalgamated using unique identifiers to provide a comprehensive analysis. The statistical analysis indicated considerable variability within the examined sample, encompassing disparities in age, height, weight, gender, engagement with medical personnel, documentation techniques, and the direction of the cardiac electrical axis. The findings underscore the complexity and diversity of the examined group, necessitating customized analytical methodologies.

### 2.3. Knowledge database discovery (KDD) process

The knowledge discovery in databases (KDD) process applied to clinical data was executed in multiple steps. During the preprocessing phase, ECG signals were purified to eliminate noise, while maintaining the integrity of cardiac cycles through sophisticated procedures. During the feature extraction phase, feature engineering was employed to identify the most pertinent properties for diagnosing atrial fibrillation (AF), hence enhancing model accuracy considerably. The hybrid predictive model was constructed utilizing multinomial logistic regression, naïve Bayes, and neutrosophic computation, facilitating the detection of concealed patterns and intricate interactions within the data. The models' robustness and dependability were confirmed via cross-validation and performance metrics, including accuracy, sensitivity, specificity, and ROC curves, thereby affirming their significance in clinical practice and cardiovascular disease management [6].

### 2.4. Hybrid predictive modeling

A hybrid model was created by integrating multinomial logistic regression, naïve Bayes, and neutrosophic computation to improve the precision and efficacy of predictions concerning atrial fibrillation (AF) [7]. The design of well-structured features was important in precisely expressing the data, resulting in significant enhancements in prediction performance in clinical settings. Advanced noise reduction techniques were employed to improve the quality and readability of ECG data. The triangle membership function from fuzzy logic was employed to distinguish the P, Q, R, S, and T waves across several ECG leads, facilitating precise detection of crucial cardiac events and enhancing

data classification accuracy [8]. This methodology markedly enhanced the model's efficacy in the identification and diagnosis of atrial fibrillation.

### 2.5. SHAP and LIME contributions

Advancements in model interpretation methodologies, including SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have markedly enhanced our comprehension of machine learning models in medical contexts, especially regarding the analysis and prediction of atrial fibrillation (AF). SHAP, developed by Arboleda Florez et al. in 2023, offers a methodology for elucidating model outputs by uniformly attributing significance to each input feature [9]. Simultaneously, LIME, introduced by Melo et al. in 2021, emphasizes the creation of local, model-agnostic explanations, rendering it applicable to intricate models such as neural networks [10]. Within the framework of the developed AF prediction tool, both methodologies are integral. SHAP allows researchers and doctors to ascertain which medical variables, such as resting heart rate or R-R interval variability, most significantly impact predictions by explicitly and transparently quantifying their contributions. LIME elucidates the rationale for a patient's classification as high-risk by the model, providing insights that are comprehensible to healthcare professionals. This interpretability is crucial for establishing confidence and facilitating the integration of machine learning systems in clinical practice.

### 2.6. Neutrosophic Values Management (SVN)

The created method incorporates Single-Valued Neutrosophic Numbers (SVN), improving analytical precision by mitigating the uncertainty present in medical data. The figures pertain to essential characteristics associated with atrial fibrillation (AF), including heart rate and R-R interval variability, employing decision-making methodologies based on neutrosophic sets. M. L. Vázquez asserts that SVNs are crucial for navigating ambiguity in intricate medical environments where measurements can fluctuate due to many influencing factors. Additionally, the program utilizes Weighted Single-Valued Neutrosophic Numbers (SVNWA) to allocate varying significance levels to each attribute, facilitating a more nuanced and precise analysis of AF-related data [11]. This method facilitates improved distinction among cardiac health conditions and enhances individualized prediction and therapy approaches. The integration of SVNWA into the system enhances clinical decision-making by evaluating the relative significance of each feature.

.

$$\mathrm{Fw}(A1, A2, \dots, An) = \langle 1 - \prod j = 1n(1 - TAj(x))wj, \prod j = 1n(IAj(x))wj, \prod j = 1n(FAj(x))wj \rangle \quad (1)$$

### 2.7. Euclidean distance

In addition to Single-Valued Neutrosophic Numbers (SVN) and their weighted variants (SVNWA), the tool incorporates Euclidean distance as a metric to assess similarity between different patient records. According to A. Sharif et al. (2019), this approach is effective in identifying common patterns and relationships within the data, significantly enhancing the accuracy of medical predictions [12]. Euclidean distance enables meaningful comparisons between historical and current cases, offering deeper insights into disease progression and supporting more informed clinical decision-making.

$$d(A_i, I) = \sqrt{\frac{1}{3}((TU_i - TU_I)^2 + (IU_i - IU_I)^2 + (FU_i - FU_I)^2)} \quad (2)$$

### 2.8. Identification of the criteria for determining the presence of atrial fibrillation

This article explores ten crucial evaluative criteria for diagnosing AF based on electrocardiographic data. In addition, an SVN number-based approach is used to assess biomarker severity and determine AF risk in patients.

**Table 1.** Criteria for determining the presence of atrial fibrillation.

| No | Evaluation criteria |
|---|---|
| $C_1$ | Extremely elevated heart rate: Monitoring of heart rates above 100 to 180 beats per minute at rest is essential for the detection of atrial fibrillation [13]. This condition may manifest with elevated rates, which may be indicative of arrhythmic episodes. |
| $C_2$ | Median R-peaks: Measurement of the median R-peaks on an ECG provides critical information about the regularity of the heart rhythm. Significant variations in the median can indicate possible irregularities in the electrical conduction of the heart [14]. This parameter is essential for identifying abnormal heart rhythm patterns. |
| $C_3$ | Intervals between heartbeats differing by more than 50 milliseconds: Measurement of RR interval variability on an electrocardiogram emerges as an essential pilot in the diagnostic navigation of atrial fibrillation [15]. Exceeding 50 milliseconds in variations becomes a beacon that reveals the irregular tides of the heart rhythm. |
| $C_4$ | Percentage of irregular R-R intervals: the percentage of irregular RR intervals relative to total intervals is a marker of heart rate variability [16]. Elevated values may indicate an increased likelihood of atrial fibrillation or other arrhythmias. |
| $C_5$ | Standard deviation of R-R intervals: The standard deviation of RR intervals reflects heart rate variability [17]. High values may indicate heart rate irregularities that require further evaluation. |
| $C_6$ | Root mean square of the mean of the squared differences of the R-R intervals: This measure provides information on the short-term variability of the heart rate, being useful for detecting atrial fibrillation and other arrhythmias [18]. |
| $C_7$ | Median Q-peaks: evaluation of median Q-peaks on an ECG may reveal abnormalities that indicate the presence of atrial fibrillation or other cardiac conditions [19]. |
| $C_8$ | Median S-peaks: The median of the S-peaks in an ECG is important for assessing the regularity of the heart rhythm [20]. Significant variations may indicate possible irregularities in ventricular repolarization. |
| $C_9$ | Median T-peaks: Evaluation of the median T-peaks on an ECG provides information about the electrical activity of the heart [21]. Abnormalities may suggest cardiac problems, including the presence of atrial fibrillation. |
| $C_{10}$ | Absence of P-peaks: The evaluation of the absence of the p-wave in an ECG is revealed as a crucial act in the accurate detection of atrial fibrillation [22]. The P wave, symbolizing atrial depolarization, constitutes a vital marker of cardiac electrical activity. |

The ten selected criteria were evaluated using a system based on Single-Valued Neutrosophic Numbers (SVN), with each criterion weighted according to its relative importance in predicting atrial fibrillation (AF). For this analysis, the weight vector was defined using the absolute values of importance: W = (**0.3, 0.3, 0.2, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0**). The following table presents the linguistic terms used to describe each criterion, along with their corresponding SVN values:

**Table 2.** Scale of linguistic terms

| Linguistic terms | SVN numbers |
|---|---|
| Extremely good (EB) | (1,0,0) |
| Very very good (MMB) | (0.9,0.1,0.1) |
| Very good (MB) | (0.8,0.15,0.20) |
| Good(B) | (0.70,0.25,0.30) |
| Medium good (MDB) | (0.60,0.35,0.40) |
| Medium(M) | (0.50,0.50,0.50) |
| Moderately bad (MDM) | (0.40,0.65,0.60) |
| Poor (MA) | (0.30,0.75,0.70) |
| Very bad (MM) | (0.20,0.85,0.80) |
| Very very bad (MMM) | (0.10,0.90,0.90) |

**Process of translating linguistic values into clinical decisions using SVNs**

1. **Assigning neutrosophic numbers to linguistic terms:**
   Each linguistic term is represented as a single-valued neutrosophic number with three components:
   **T** (truth-membership): degree of truth (e.g., 0.9 for "Extremely good")
   **I** (indeterminacy-membership): degree of indeterminacy or uncertainty
   **F** (falsity-membership): degree of falsity.

   **For example:**
   "Extremely good" → SVN = (T = 0.9, I = 0.05, F = 0.05)
   "Good" → SVN = (0.8, 0.1, 0.1)
   "Fair" → SVN = (0.6, 0.2, 0.2).

2. **Multicriteria evaluation:**
   Clinical criteria relevant to the diagnosis of AF are defined (e.g., avg_hr, sdnn, rmssd, etc.), and each is evaluated using linguistic terms. These terms are translated into their corresponding SVNs.

3. **Weighting using SVNWA (Single-Valued Neutrosophic Weighted Averaging):**
   Weights are applied to each criterion according to their relative importance (e.g., 0.3, 0.2, etc.).
   Neutrosophic aggregation is then performed for each patient using the SVNWA formula, which combines the weighted T, I, and F values.

4. **Calculation of similarity to the ideal:**
   - The neutrosophic Euclidean distance between the patient's SVN vector and an ideal vector (e.g., one with T = 1, I = 0, F = 0) is calculated.
   - The closer a patient is to the ideal vector, the greater their similarity to the optimal health state.

5. **clinical decision-making:**
   - The resulting similarity score is interpreted as an indicator of the patient's risk or clinical status.

- For example, an aggregate result with a high T value and low I and F values is translated as a positive or controlled clinical status.
- Conversely, if the aggregate result shows a low T value and a high F value, it is interpreted as an increased risk of AF, which guides the physician to more intensive intervention or monitoring.

**Practical example of the process:**
Suppose a patient has characteristics assessed as:
avg_hr: "Good" → (0.8, 0.1, 0.1)
sdnn: "Fair" → (0.6, 0.2, 0.2)
With weights: W = (0.6, 0.4)
The SVNWA formula is applied:

$$SVNWA = (\sum w_i T_i, \sum w_i I_i, \sum w_i F_i) \quad (3)$$

T aggregate = 0.6 × 0.8 + 0.4 × 0.6 = 0.72
I aggregate = 0.6 × 0.1 + 0.4 × 0.2 = 0.14
F aggregate = 0.6 × 0.1 + 0.4 × 0.2 = 0.14

This result (0.72, 0.14, 0.14) indicates a relatively favorable but suboptimal state, which can be interpreted as a low risk or controlled condition, useful for making individualized clinical decisions.

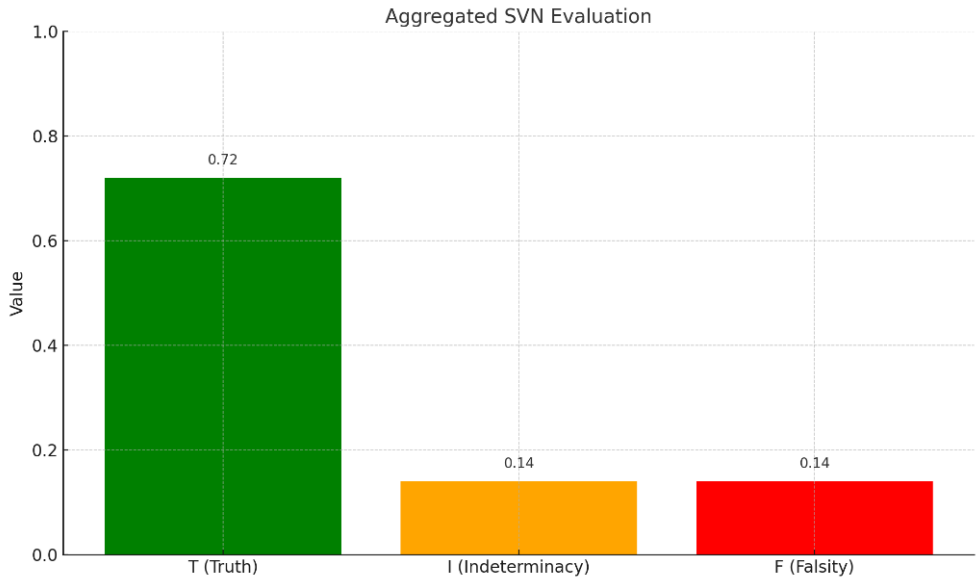


**Figure 1.** The chart shows that the evaluation outcome has a high degree of truth (0.72), with low levels of indeterminacy (0.14) and falsity (0.14), indicating strong support for a favorable clinical decision.

Once the evaluation framework was established and the specific criteria were defined, a thorough collection of relevant data was carried out. The table below displays the assessment of ten criteria (c1 to c10) used to determine the presence of atrial fibrillation (AF) in three different patients, identified as X1, X2, and X3.

**Table 3.** Resultados de la recogida de información

| | x1 | x2 | x3 |
|---|---|---|---|
| e1 | EM | MMM | MDM |
| e2 | MMM | MM | MDM |
| e3 | M | MA | MDM |
| e4 | M | MA | MDM |
| e5 | M | MDM | MDM |
| e6 | M | M | MDM |
| e7 | M | M | M |
| e8 | M | M | M |
| e9 | M | M | M |
| e10 | M | M | M |

The clinical characteristics of the three evaluated patients are detailed below.

**Patient 1** shows an average heart rate (avg_hr) of 375, nn50 of 9, P wave median (p_median) of 0.00, pnn50 of 75, q_median of 0.31, r_median of 0.68, rmssd of 0.13, s_median of 0.10, sdnn of 0.09, and t_median of 0.44.

**Patient 2** presents with an avg_hr of 300, nn50 of 11, p_median of 0.59, pnn50 of 73, q_median of 0.58, r_median of 0.96, rmssd of 0.22, s_median of 0.51, sdnn of 0.15, and t_median of 0.58. Finally,

**Patient 3** has an avg_hr of 286, nn50 of 8, p_median of 0.00, pnn50 of 57, q_median of 0.43, r_median of 0.82, rmssd of 0.30, s_median of 0.04, sdnn of 0.18, and t_median of 0.36.

To consolidate the assessments of the different decision makers into a single representative value, the SVNWA aggregation operator was used. This operator allows the uncertainty and indeterminacy of the individual evaluations to be managed by appropriately weighting the different characteristics of each patient. The results of the aggregation, the corresponding scoring and the ranking of the patients based on these evaluations are presented below.

**Table 4.** Evaluation results

| Patient | Aggregation |
|---|---|
| $\mathbf{X_1}$ | (0.05, 0.95, 0.95) |
| $\mathbf{X_2}$ | (0.22, 0.81, 0.78) |
| $\mathbf{X_3}$ | (0.40, 0.65, 0.60) |

As a reference, we have identified the ideal alternative represented by the vector:

$$E^{+} = \textbf{(EM, MMM, MA, MA, MDM, MDM, MDM, MDM, MDM, MDM)}$$

This vector defines the optimal or desirable attributes in our data, setting a standard that we aspire to achieve in the evaluation of individual profiles. The Euclidean distances calculated for the patients evaluated are presented below:

**Table 5.** Calculation of distance

| Patient | Euclidean distance |
|---|---|
| $\mathbf{X_1}$ | 0.42 |
| $\mathbf{X_2}$ | 0.3 |
| $\mathbf{X_3}$ | 0.55 |

These distances reflect the proximity or divergence of each profile with respect to our ideal represented by $E^+$ El The system uses these metrics to guide clinical decisions, providing a detailed and objective assessment of each case.

## 3. Results

### 3.1. Descriptive Analysis

Initial descriptive analysis of the dataset provided a comprehensive view of the measured cardiac characteristics, which include:

- Average heart rate (avg_hr)
- Median of R-R intervals (r_median)
- Number of successive differences of R-R intervals greater than 50 ms (nn50)
- Percentage of nn50 (pnn50)
- Standard deviation of the R-R intervals (sdnn)
- Root mean square of the mean of the squared differences of the R-R intervals (rmssd)
- Median of peak Q (q_median)
- Median of peak S (s_median)
- Median of peak T (t_median)
- Mediana del pico P (p_median)
- Rhythm classification (ritmi)

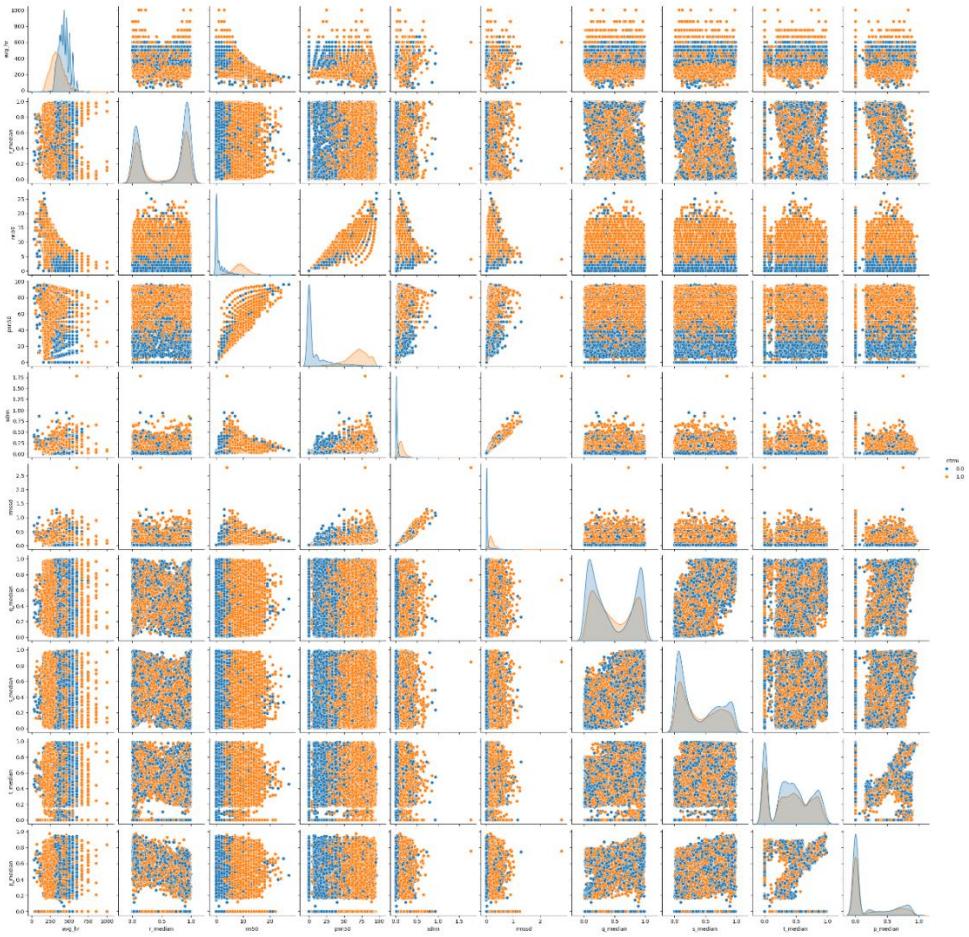

**Figure 2.** Matrix of scatter plots and variable distributions.

The figure displays a matrix of plots illustrating the distribution and dispersion of several variables in the dataset. Each cell contains a scatter plot showing the relationship between two variables, while the diagonal plots depict their individual distributions. The categorical variable "ritmi" is used to assign point colors: blue represents normal sinus rhythm (0), and orange indicates atrial fibrillation (1). This visualization served as a foundational tool for subsequent analysis, aiding in the detection of significant patterns and irregularities in cardiac activity.

**I. Diagonal distributions:**

The plots along the diagonal display the distribution of each individual variable. Some variables show skewed or bimodal distributions, reflecting internal differences among patient subgroups. This suggests that these variables may carry useful signals for diagnostic analysis.

**II. Relationships between variables:**

The off-diagonal cells reveal bivariate relationships between variables. In several pairings, the blue and orange dots tend to form distinct clusters, indicating that certain variable combinations may help distinguish between normal sinus rhythm and atrial fibrillation. This visual separation is valuable for guiding feature selection in predictive models.

**III. Density and dispersion:**

Variables display varying levels of dispersion—some are tightly clustered while others are more spread out. This variability can influence the importance of each variable within a model, as it affects the stability of their values. More stable, less dispersed variables tend to be more reliable in clinical classification processes.

**IV. Potentially relevant variables:**

Variables such as ***avg_hr***, ***rmssd***, ***sdnn***, and ***pnn50*** show clearer separation between classes, making them strong candidates as key predictors of atrial fibrillation. This distinction suggests that these physiological features may reflect significant cardiac abnormalities. Including them in models can enhance diagnostic sensitivity.

**V. Visual overlap:**

While some variables show clear separation, many others exhibit substantial overlap between the orange and blue dots. This indicates that no single variable is fully decisive, supporting the use of multivariate or hybrid approaches. Combining multiple variables is essential for improving analytical accuracy.

### 3.2. Multinomial Logistic Regression

Multinomial logistic regression is a technique used to classify observations into more than two categories, making it particularly valuable in clinical settings where cardiac conditions may vary. The model is configured with key parameters, such as the number of classes, the optimization method (e.g., newton-cg), and a maximum number of iterations to ensure convergence. Once defined, it is trained using a feature matrix representing relevant physiological variables and a label vector containing the corresponding classes. This process enables the model to identify statistical patterns that link clinical indicators with the presence or absence of atrial fibrillation (AF) or other related conditions.

Once trained, the model is used to make predictions on new data. For each observation in the test set, the probability of belonging to each class is calculated using the multinomial logistic function.

The class with the highest probability is then assigned as the predicted label. This ensures that each patient is classified based on the most likely condition, taking all clinical variables into account. The predictions are stored for subsequent analysis or validation, allowing researchers to assess the model's accuracy on real-world data. Overall, the algorithm provides a robust decision-support tool grounded in multivariate clinical data.

Mathematical Model of Multinomial Logistic Regression

Below is the mathematical model of Multinomial Logistic Regression, also known as SoftMax Regression, formulated for a classification problem with multiple classes (in this case, to predict atrial fibrillation or other cardiac conditions):

Suppose you have a set of possible classes
C= {1, 2, ..., K},
and a feature vector $\mathbf{x} \in \mathbf{R}^n$.
Probability function (softmax):

$$P(y = k/x) = \frac{\exp(\beta_k x)}{\sum_{j=1}^{k} \exp(\beta_j x)} \quad (4)$$

**Where:**

- $\mathbf{B}_k \in \mathbf{R}^n$ is the coefficient vector associated with class **k**,
- $\mathbf{B}^t{}_k \mathbf{X}$ is the dot product between the coefficients and the input features,
- **exp (·)** is the exponential function,
- The denominator ensures that the probabilities sum to 1 (normalization).

**Class prediction:**

$$\hat{y} = \underset{k \varepsilon C}{\arg max}\ P(y = k/x) \quad (5)$$

### 3.3 Naive Bayes Gaussiano

The Gaussian Naive Bayes model operates under the premise of variable independence and posits that each characteristic adheres to a normal distribution within each class. In the training phase, the model computes the mean and standard deviation of each variable for every class utilizing the training dataset. These factors are utilized to formulate Gaussian density functions that assess the probability of detecting particular feature values for a designated class. This computational simplicity renders it an effective and rapid model, particularly advantageous for high-dimensional clinical datasets.

Once trained, the model evaluates new observations by computing the probability of class membership, taking into account all available clinical variables. It then assigns the class with the highest probability as the final prediction. Although it assumes independence among features—which may not always hold in medical contexts—its performance is often competitive in practice, particularly for initial classification tasks. This technique is useful for identifying general patterns in the data and supporting early clinical decision-making through a solid and easily interpretable probabilistic approach.

***Mathematical Model of Gaussian Naive Bayes***

Given a feature vector $X = (x_1, x_2, \ldots, x_n)$ and a set of possible classes $C = \{c_1, c_2, \ldots, c_k\}$, the goal is to predict the most probable class for a given observation.

**Bayes' Theorem:**

$$P(C_k/x) = \frac{P(C_k)*P(x/C_x)}{P(x)} \quad (6)$$

Since $\boldsymbol{P(x)}$ is constant across all classes, classification is based on:

$$P(C_k/x) \propto P(C_k) * P(x/C_x)$$

Conditional Independence Assumption (Naive):

$$P(x \Big/ c_k) = \frac{1}{\sqrt{2\,\pi\,\sigma_{jk}^2 \exp\left(-\frac{\left(x_j-\mu_{jk}\right)^2}{2\,\sigma_{jk}^2}\right)}} \quad (7)$$

**Where:**

$\mu_{jk}$ is the mean of feature $\boldsymbol{x_k}$ for class **ck,**

$\sigma_{jk}^2$ is the corresponding variance**.**

**Class Prediction Rule:**

$$\hat{\boldsymbol{y}} = \underset{c_k \in C}{\arg max}\ P(c_k) * \prod_{j=1}^{n} P(x_j/c_k) \quad (8)$$

**Interpretation:**

This model is simple yet powerful. It works well when feature distributions are approximately normal and is computationally efficient even with high-dimensional data. While the independence assumption rarely holds perfectly in practice, Gaussian Naive Bayes often performs remarkably well in classification tasks.

**Confusion Matrix Analysis for Model Performance Comparison**

The confusion matrices offer a comprehensive visual comparison of the classification efficacy of Multinomial Logistic Regression and Gaussian Naive Bayes in identifying atrial fibrillation (AF). These matrices encapsulate model predictions for two categories—normal sinus rhythm and atrial fibrillation—utilizing criteria like precision, recall, and F1-score. Multinomial Logistic Regression exhibited somewhat improved class equilibrium and reduced misclassifications, especially among healthy patients. Concurrently, Gaussian Naive Bayes demonstrated robust recall for AF cases, underscoring its sensitivity. Together, the visual analysis emphasizes each model's strengths and supports their potential use in clinical decision support systems.

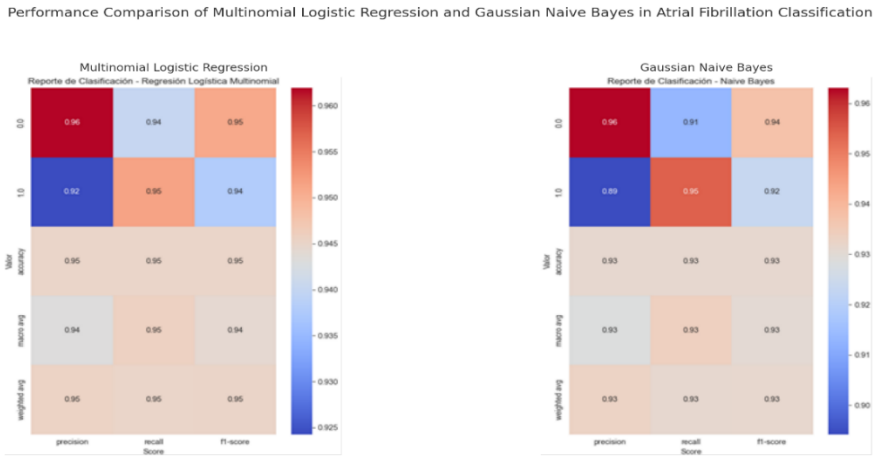


**Figure 3.** Ranking report for the multinomial and naive Bayes logistic regression models; (a) The image shows a ranking report evaluating the performance of the multinomial logistic regression model; (b) The image shows a ranking report comparing the performance of the naive Bayes model.

In both cases, key metrics such as accuracy, recall, F1-score, and support are included for each class. These metrics are fundamental to evaluate the accuracy and efficiency of each model in classifying data.

**Multinomial Logistic Regression**: Este modelo mostró un desempeño destacado con una precisión del 94.51%. La matriz de confusión reveló una alta capacidad de clasificación correcta para ambas clases, fibrilación auricular y ritmo sinusal normal. El reporte de clasificación indicó una alta precisión, recall y f1-score para ambas clases, con una precisión global del modelo del 95%. La curva ROC, con un área bajo la curva (AUC) de 0.98, subraya la excelente capacidad del modelo para distinguir entre las dos clases. El Coeficiente de Correlación de Matthews (MCC) de 0.8889 refuerza la robustez del modelo en términos de equilibrio entre las clases positivas y negativas.

El modelo **Naive Bayes Gaussiano** también demostró un rendimiento competente con una precisión del 93.11%. La matriz de confusión y el reporte de clasificación indicaron una alta precisión y recall para ambas clases, aunque ligeramente inferior al modelo de Regresión Logística Multinomial. La curva ROC con un AUC de 0.97 destaca una capacidad de discriminación efectiva. El MCC de 0.8624 sugiere un rendimiento equilibrado, aunque ligeramente menor comparado con la Regresión Logística Multinomial.

### 3.4. Model Evaluation.

The performance of both models was thoroughly compared using evaluation metrics such as precision, recall, F1-score, ROC curve AUC, and Matthews Correlation Coefficient (MCC). Multinomial Logistic Regression achieved slightly better overall accuracy and class balance. However, Gaussian Naive Bayes also proved to be an effective model, especially for handling data that follow a normal distribution.

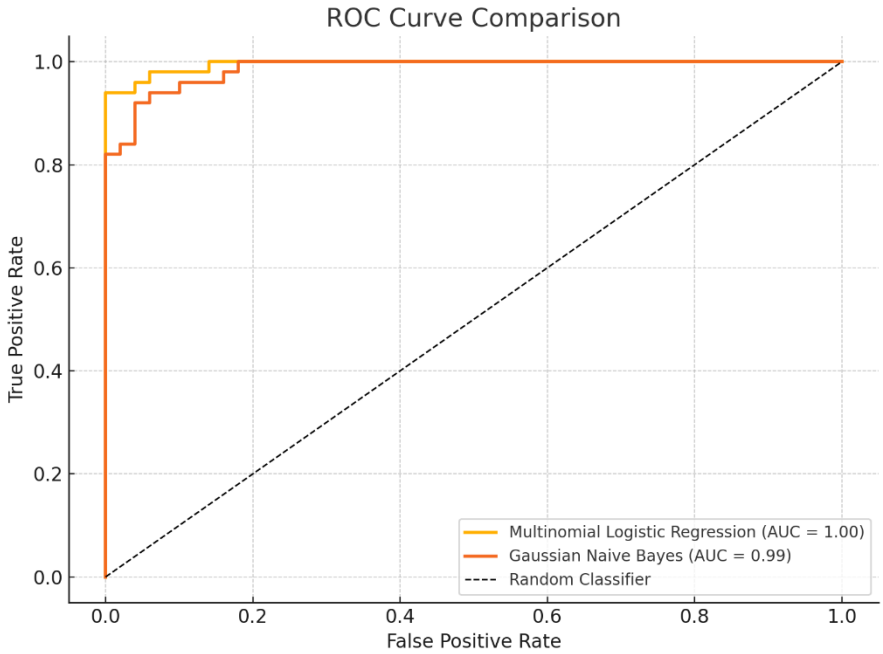


**Figure 4.** ROC curves for the multinomial and Naive Bayes logistic regression models; (c) The image shows the ROC curve evaluating the performance of the multinomial logistic regression model; (d) The image shows the ROC curve comparing the performance of the Naive Bayes model. In both cases, key metrics such as the area under the curve (AUC) are included, which is fundamental to evaluate the ability of each model to distinguish between classes

The ROC curve comparison demonstrates that both models—Multinomial Logistic Regression and Gaussian Naive Bayes—exhibit exceptional classification performance in detecting atrial fibrillation, with areas under the curve (AUC) around 1.00 and 0.99, respectively. The Logistic Regression curve consistently surpasses that of Naive Bayes at most thresholds, demonstrating a superior balance of sensitivity and specificity. Although both models are effective, the superior AUC of the Logistic Regression model indicates it offers more precise and dependable differentiation between normal sinus rhythm and atrial fibrillation, rendering it a more favorable option in clinical decision-making scenarios.

## 4. Applications.

In the extensive realm of cardiac data, data integrity commences with meticulous preparation. This entails addressing absent values and standardizing ranges to uncover concealed patterns in the heart's electrical activity, facilitating a more profound and significant comprehension of atrial fibrillation (AF). Advanced methodologies, including noise reduction and peak detection, facilitated accurate analysis and the early discovery of potential anomalies. Structuring the processed data into tabular formats improved clarity and accessibility for subsequent study and simulations.

Moreover, employing analytical instruments like Euclidean distance and similarity metrics to an ideal reference offers an objective and quantifiable method for evaluating the degree to which individual profiles align with optimal patterns. The incorporation of machine learning techniques and predictive algorithms enhanced the capacity to identify latent trends in cardiac signals and increased the precision of atrial fibrillation risk assessment, facilitating the formulation of individualized treatment regimens tailored to each patient's unique profile.

## 5. Conclusion

The creation of a novel intelligent platform seeks to surpass conventional methods by addressing the difficulty of identifying biomarkers for atrial fibrillation (AF). This development combines neutrosophic statistical techniques with machine learning, utilized on anonymised ECG data. The system's technological core comprises a meticulously organized framework utilizing React.js, React Router, Flask, Node.js, and CouchDB, rather than merely a collection of tools. Each element facilitates a cohesive user experience and permits the adaptive handling of cardiac data in accordance with contemporary cardiovascular diagnostics requirements.

This method enhances the accuracy and efficacy of atrial fibrillation identification while enabling the lucid interpretation of intricate data, hence bolstering evidence-based clinical decision-making, and progressing cardiovascular diagnostics in practical medical environments. This innovation signifies a significant advancement in the realm of digital cardiology. The platform establishes a robust foundation for future research by incorporating advanced tools for data analysis and processing, so creating new opportunities for more effective treatments of atrial fibrillation and other cardiac arrhythmias. Its versatility and scalability facilitate its use in various therapeutic settings, potentially revolutionizing medical practice by enabling more personalized and accurate therapies adapted to individual patient needs.

## References

[1] Fácila Rubio, L.; Lozano-Granero, C.; Vidal-Pérez, R.; Barrios, V.; Freixa-Pamias, R. Nuevas tecnologías para el diagnóstico, tratamiento y seguimiento de las enfermedades cardiovasculares. Rev. Esp. Cardiol. 2023, Jul. https://doi.org/10.1016/J.RECESP.2023.07.011.

[2] Cevallos-Torres, L.; Núñez-Gaibor, J.; Leyva-Vasquez, M.; Gómez-Rodríguez, V.; Parrales-Bravo, F.; Hechavarría-Hernández, J. NCC: Neutrosophic Control Charts, a didactic way to detect Cardiac Arrhythmias from reading Electrocardiograms. Neutrosophic Sets Syst. 2024, 74(1), 40.

[3] Parrales-Bravo, F.; Cevallos-Torres, L.; Vasquez-Cevallos, L.; Bazurto-Blacio, Z.; Minda-Gilces, D.; Gómez-Rodríguez, V.; García-Arias, P. Current Situation of the Use of Neutrosophy to Improve Medical Decision Making and Preventive Healthcare in Ecuador. Neutrosophic Sets Syst. 2025, 84, 110–122.

[4] Gamal, M.; Zaied, A.N.H.; Rushdy, E. Ensemble Classifiers for Acute Leukemia Classification Using Microarray Gene Expression Data under Uncertainty. Neutrosophic Sets Syst. 2022, 49, 164–183. https://doi.org/10.5281/zenodo.6426383.

[5] Wagner, P.; et al. PTB-XL, a large publicly available electrocardiography dataset. Sci. Data 2020, 7(1), 1–15. https://doi.org/10.1038/s41597-020-0495-6.

[6] Corona Organiche, E.; Jorge, A.; Alfaro, J.; Barrera, G.C. Principales Metodologías en el Desarrollo de Proyectos de Minería de Datos. TecnoCultura 2020, 51–51. https://tecnocultura.org/index.php/Tecnocultura/article/view/9.

[7] Parrales-Bravo, F.; Caicedo-Quiroz, R.; Barzola-Monteses, J.; Cevallos-Torres, L. Prediction of emergency room arrivals of patients with preeclampsia disease using artificial neural network model. In 2024 IEEE 4th Int. Conf. Electron. Commun., IoT and Big Data (ICEIB) 2024, 34–39. IEEE.

[8] Hampton, J. The ECG Made Easy, 9th ed.; 2019; pp. 1–269.

[9] Arboleda-Florez, M.; Castro-Zuluaga, C. Interpreting direct sales' demand forecasts using SHAP values. Production 2023, 33. https://doi.org/10.1590/0103-6513.20220035.

[10] Melo, R.O.; de Freitas, A.C.; Francisco, E.R.; Motokane, M.T. Impact of socioeconomic variables on Enem performance: a spatial and sociologic analysis. Rev. Adm. Publica 2021, 55(6), 1271–1294. https://doi.org/10.1590/0034-761220200843.

[11] Paucar, E.C.; Ramírez, J.L.M.; Vela, R.C.; Flores, J.L.A.; Pezo, V.L.; Águila, J.R.R.; Paredes, V.R.R. Neutrosophic Assessment of Legal Certainty in the Digital Transformation of Logistics in Callao: A Comparative Study Based on SVNN and the Neutrosophic Analytic Hierarchy Process (AHP). Neutrosophic Sets Syst. 2025, 84, 742–754.

[12] Sharif, A.; et al. Intelligent human action recognition: A framework of optimal features selection based on Euclidean Distance and Strong Correlation. Control Eng. Appl. Informatics 2019, 21(3), 3–11.

[13] LeWine, M.H.E. Cardiac Arrhythmias - Harvard Health. Harvard Health Publ. Accessed Jul. 11, 2024. https://www.health.harvard.edu/a_to_z/cardiac-arrhythmias-a-to-z.

[14] Bae, T.W.; Lee, S.H.; Kwon, K.K. An adaptive median filter based on sampling rate for R-peak detection and major-arrhythmia analysis. Sensors 2020, 20(21), 1–21. https://doi.org/10.3390/s20216144.

[15] Khan, A.A.; Junejo, R.T.; Thomas, G.N.; Fisher, J.P.; Lip, G.Y.H. Heart rate variability in patients with atrial fibrillation and hypertension. Eur. J. Clin. Invest. 2021, 51(1), 1–13. https://doi.org/10.1111/eci.13361.

[16] Corino, V.D.A.; et al. Association between atrial fibrillatory rate and heart rate variability in patients with atrial fibrillation and congestive heart failure. Ann. Noninvasive Electrocardiol. 2013, 18(1), 41–50. https://doi.org/10.1111/anec.12019.

[17] Jin, M.N.; et al. Long-Term Outcome of Additional Superior Vena Cava to Septal Linear Ablation in Catheter Ablation of Atrial Fibrillation. J. Am. Heart Assoc. 2019, 8(22). https://doi.org/10.1161/JAHA.119.013985.

[18] Shen, M.; Zhang, L.; Luo, X.; Xu, J. Atrial fibrillation detection algorithm based on manual extraction features and automatic extraction features. IOP Conf. Ser. Earth Environ. Sci. 2020, 428(1). https://doi.org/10.1088/1755-1315/428/1/012050.

[19] Eliasson, H. Robust Extraction of Respiratory Signals from ECG. Signal. 2024, June.

[20] Rogovoy, N.M.; et al. Hemodialysis Procedure–Associated Autonomic Imbalance and Cardiac Arrhythmias: Insights From Continuous 14-Day ECG Monitoring. J. Am. Heart Assoc. 2019, 8(19). https://doi.org/10.1161/JAHA.119.013748.

[21] Bin Sinal, M.S. Quick Identification of Arrhythmia Symptoms Using Empirical Approach in Long Sequence of Heart Cycles. Shibaura Inst. Technol. 2019, Rsj, 2002–2004.

[22] Waugh, J.L.S.; Patel, R.; Ju, Y.; Patel, A.B.; Rusin, C.G.; Jain, P.N. A novel automated junctional ectopic tachycardia detection tool for children with congenital heart disease. Hear. Rhythm O2 2022, 3(3), 302–310. https://doi.org/10.1016/j.hroo.2022.02.014.

Received: May 30, 2025. Accepted: July 09, 2025.