www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(1), Jan-April 2026, 275-285. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2026.v17i1.1145

Identification of Cardiac Arrhythmia by Selection of Relevant Variables Using Genetic Algorithms

[1]Santiago Arias-García, [2] José Hernández-Torruco, [3] Betania Hernández-Ocaña, [4] Oscar Chávez-Bosquez
Universidad Juárez Autónoma de Tabasco
2 email:jose.hernandezt@ujat.mx

Abstract. This article presents the computational identification of cardiac arrhythmia from electrocardiogram (ECG) signal recordings to facilitate timely diagnosis and clinical management. The cardiac arrhythmia dataset from the public UCI repository was used, comprising 279 features and 452 classified cases. Several variable selection algorithms were applied, including filter methods such as OneR, Chi-square, information gain, symmetric uncertainty, gain ratio, CFS, and consistency, as well as a metaheuristic approach based on the genetic algorithm. The variables identified through the filter methods were subsequently used as inputs for the OneR, PART, Rpart, JRip, C4.5, SVMlin, KNN, and random forest classifiers. The results indicate two subsets of particular interest: 37 relevant variables achieving an average balanced accuracy of 82.93% using the Random Forest classifier in combination with the CFS filter method, and 12 relevant variables yielding an average balanced accuracy of 82.56% when the Random Forest classifier is combined with the genetic algorithm. These outcomes were obtained without the application of any data balancing method.
Keywords: arrhythmia; classification; relevant variables; ECG; genetic algorithm.

Article Info
Received June 3, 2025
Accepted July 2, 2025

1 Introduction

Technological advances have significantly increased the tools available to detect and predict diseases. Healthcare is a crucial area that scientists are addressing through machine learning, a field that focuses on critical feature extraction and predictive analytics. This has created models that can help physicians deliver timely treatment and improve the quality of medical care.

Cardiovascular diseases remain a significant problem in all countries because they rank among the leading causes of mortality worldwide(Jangra et al., 2021; Morganroth, 1983). Electrocardiography (ECG) is a valuable tool for detecting heart disease by collecting information about the electrical signals emitted by the heart. Artificial Intelligence, in the form of Machine Learning, can be used to create predictive models capable of detecting cardiovascular disease using historical patient records, which can help physicians deliver timely treatment(Chakraborty et al., 2022; Leng et al., 2015).

Although models exist to classify cardiovascular diseases, such as cardiac arrhythmia, they require many variables, are computationally expensive to predict. However, it is possible to generate feasible models that perform this prediction with a reduced number of features using relevant variable selection techniques, such as filtering methods and metaheuristic algorithms that reduce the size of the dataset by discarding irrelevant information.

This study conducted experiments with binary classification algorithms to determine normal and abnormal cases of cardiac arrhythmia using a real dataset from the public repository UCI (Guvenir H. & Quinlan, 1998). Relevant variable selection techniques, using filter methods and metaheuristics called genetic algorithms, were applied to reduce the number of features and create more efficient predictive models. The results showed a significant average balanced accuracy in detecting cardiac arrhythmia using classification algorithms.

Some studies on cardiac arrhythmia detection in binary-multi classes using the same dataset are the following. Elsayad (Elsayad, 2009) utilized principal component analysis (PCA) to generate a set of 100 features and by applying LVQ2.1, obtained an accuracy of 74.12%. Jadhav et al. (Jadhav et al., 2010) designed three types of artificial neural network models: a multilayer perceptron

(MLP), a generalized feed-forward neural network (GFFNN), and a modular neural network (MNN), obtaining accuracies of 86.67%, 82.35%, and 82.22%, respectively, using 279 features. Shensheng Xu et al. (Xu et al., 2017)achieved 82.96% accuracy using 236 features selected using Fisher's discriminant index (FDR) and PCA. Ayar et al. (Ayar & Sabamoniri, 2018) employed genetic algorithms (GA) and decision trees; their proposal identified 61 features with an accuracy of 86.96%. Kadam et al. (Kadam et al., 2020) used soft margin SVM and elitist GA to find 92 features and achieved 87.83% accuracy with 10-fold cross-validation. C. K. Roopa et al. proposed an unsupervised method for classifying cardiac arrhythmias using ECG data from the UCI dataset. They employed LDA, RLPI, and PCA for feature selection and clustered data using the RSKFCM method, achieving an accuracy of 74.91% with PCA, 89.24% with LDA, and 96.41% with RLPI(Roopa et al., 2018). In 2019, Saroj Kumar et al. used neural networks (RNN, feed-forward backpropagation networks, and radial basis function networks), achieving a maximum accuracy of 83.1% with RNN (Pandey & Janghel, 2019). Parven et al. employed SVM and K-NN, achieving 97.87% and 96.85% accuracy, respectively (Parveen et al., 2021). In 2023, Bashar et al. applied artificial neural networks, achieving an accuracy of 92.47%, sensitivity of 93.03%, and specificity of 92.03%(Al-Saffar et al., 2023).

This paper is structured as follows: Section 2 discusses the dataset, evaluation metrics, classifiers, and methods for selecting relevant variables. Section 3 presents the experimental design and discusses the obtained results. Section 4 presents the study's conclusions and prospects for future work.

2 Dataset and methods

2.1 Dataset

The UCI cardiac arrhythmia dataset contains patient information with ECG signals classified for differentiating between the presence and absence of cardiac arrhythmia. This set comprises 279 variables and 452 records, classified into 16 arrhythmia types. To improve prediction, we converted the dataset to a binary class because of the imbalance in the data concerning various arrhythmia types. Class 01 indicates a normal (healthy) ECG, and class 02 indicates cardiac arrhythmia (diseased).

Table 1 shows the original distribution of arrhythmia classes and the number of instances, and Table 2 presents the binary class distribution of the dataset. In a previous study, preprocessing of the dataset was performed (Arias García et al., 2021). Missing values were imputed using the mean, and one variable with more than 83% missing data was removed, leaving 278 predictor features.

Table 1. Original distribution of the classes.

# Class	Class Description	# instances
01	Normal	245
02	Ischemic changes (coronary artery disease)	44
03	Old anterior myocardial infarction	15
04	Older inferior myocardial infarction	15
05	Sinus tachycardia	13
06	Sinus bradycardia	25
07	Premature ventricular contraction (PVC)	3
08	Premature supraventricular contraction	2
09	Left bundle branch block	9
10	Right bundle branch block	50
11	1st degree of atrial ventricular block	0
12	2nddegree of AV block	0
13	3rd degree of AV block	0
14	Left ventricular hypertrophy	4
15	Atrial fibrillation or flutter	5
16	Others	22

Table 2. Final distribution of the classes.

# Class	Class Description	# instances
01	Normal	245
02	Cardiac arrhythmia	207

2.2 Performance measures

This study employs performance evaluation metrics from machine learning, such as accuracy, balanced accuracy, sensitivity, specificity, and ROC curve. Among all these, the primary metric we considered was balanced accuracy, which was defined as the average between specificity (healthy patients) and sensitivity (patients with arrhythmia).

Balanced accuracy refers to the arithmetic mean of sensitivity and specificity, which are the correctly identified negative and positive values, respectively (Han et al., 2011). In other words, the model correctly classified the average proportion of healthy individuals and patients with cardiac arrhythmia. Balanced accuracy is calculated as follows:

$$Balanced\ Accuracy\ = (\frac{TN}{TN + FP} + \frac{TP}{TP + FN})/2 \quad (1)$$

Where:
TP: true positive, FP: false positive; TN: true negative, and FN: false negative.

2.3 Classifiers

The classifiers used in this study were OneR, PART, Rpart, JRip, C4.5, Support Vector Machine, K-Nearest Neighbor, and the combined method Random Forest. These classifiers are described as follows: OneR is an algorithm that builds a classifier based on a single rule (1-R). Although simple, it is surprisingly accurate. In technical terms, this algorithm focuses on a single feature, i.e., it selects a specific feature for which one or more decision rules are defined(Holte, 1993; Nevill-Manning et al., 1995).

PART, on the other hand, builds a partial decision tree based on C4.5 and, at each iteration, generates the best possible rule following the approach developed by Frank and Witten(Frank & Witten, 1998).

Rpart implements decision trees using recursive partitioning. This algorithm is simple to interpret, and its visualization highlights the most relevant features employed by the classification model(Terry Therneau et al., 2022).

JRip is a rule-based classifier that employs a repeated incremental pruning technique known as RIPPER (Repeated Incremental Pruning to Produce Error Reduction) to minimize errors(Hornik, 2012). The rules generated by this algorithm have the following structure,as shown in equation 2.

if (attribute1<relational operator >value1<logical operator>
attribute2<relational operator>value2<⋯>), then (2)
value-decision

C4.5 is a rule-based algorithm that constructs a decision tree where patterns are represented as sets of "if-then" rules, thus facilitating the classification of new cases (Salzberg, 1994). In our study, pruning techniques were applied to the generated decision trees.

A support vector machine (SVM) learns to distinguish between two distinct classes by creating a decision surface from the input points. Using the information provided by the support vectors, the SVM can establish a decision boundary that delimits the data domain(BETANCOURT, 2005). In this study, we employ the linear kernel (SVMLin) by adjusting different values of parameter C to optimize the model performance. To implement the SVMLin classifier, we used the "caret" package in R(Kuhn, 2008).

K-nearest neighbor (k-NN) classification is a widely used technique, especially when attribute values are continuous. The algorithm assigns a class to a new instance by considering the majority class among its *k*nearest neighbors. In this study, we implement this algorithm using the kknn package (Hechenbichler et al., 2016).

The Random Forest algorithm (Breiman, 2001) is a prediction technique based on a set of CART trees generated using the Bootstrap method. The algorithm creates multiple Bootstrap samples from n training data points and*m* predictor variables by selecting n data points with replacements from the original data set. Then, a CART tree is trained on each Bootstrap sample using m predictors randomly chosen from the original set. Random forests consist of multiple tree-based predictors, with each tree relying on values from a randomly sampled vector independently generated with the same distribution across all trees in the forest.

2.4 Selection of relevant variables

Datasets have both essential and non-essential features. Non-essential features include irrelevant information that can slow down classification algorithms and decrease their accuracy. Feature selection methods reduce the dimensionality of a dataset, retaining only the most relevant features to enhance the performance of classification models (Dash & Liu, 2000). This study focuses on seven filtering methods from the "FSelector" package: OneR, Chi-Square, Information Gain, Symmetric Uncertainty, Gain Ratio, CFS, and Consistency.

OneR is a filter that assigns weights to attributes using simple rules involving only one attribute in the condition. It generates a basic rule for each attribute and calculates its error rate.

The Chi-Square filter assigns weights to attributes by applying the chi-square test, which evaluates the statistical relationship between each feature and the dataset's class(Zheng et al., 2004).

Information Gain computes attribute weights based on their correlation with the continuous class attribute, measuring a feature's predictive power for the class.

Symmetrical Uncertainty is an entropy-based filter, similar to Information Gain and Gain Ratio. It assigns weights to attributes by assessing their correlation with the class.

CFS is a correlation-based feature selection algorithm that assesses feature subsets through heuristic evaluation. It removes redundant features that show correlation with other features(Hall, 1999).

Consistency is an algorithm that identifies a subset of attributes by means of a consistency measure applicable to both discrete and continuous data. The selected subset of features is intended to be the smallest possible (Dash & Liu, 2003).

Genetic algorithms (GA) are an optimization technique based on heuristic algorithms and population search, emulating natural evolution. GA operations consist of iterative procedures that manipulate a population of chromosomes (possible solutions) to generate a new population through genetic functions such as crossover and mutation. This approach resembles the evolutionary principles of reproduction, genetic recombination, and survival described by Charles Darwin. GAs are adaptable and efficient methods for selecting relevant variables (Babatunde et al., 2014).

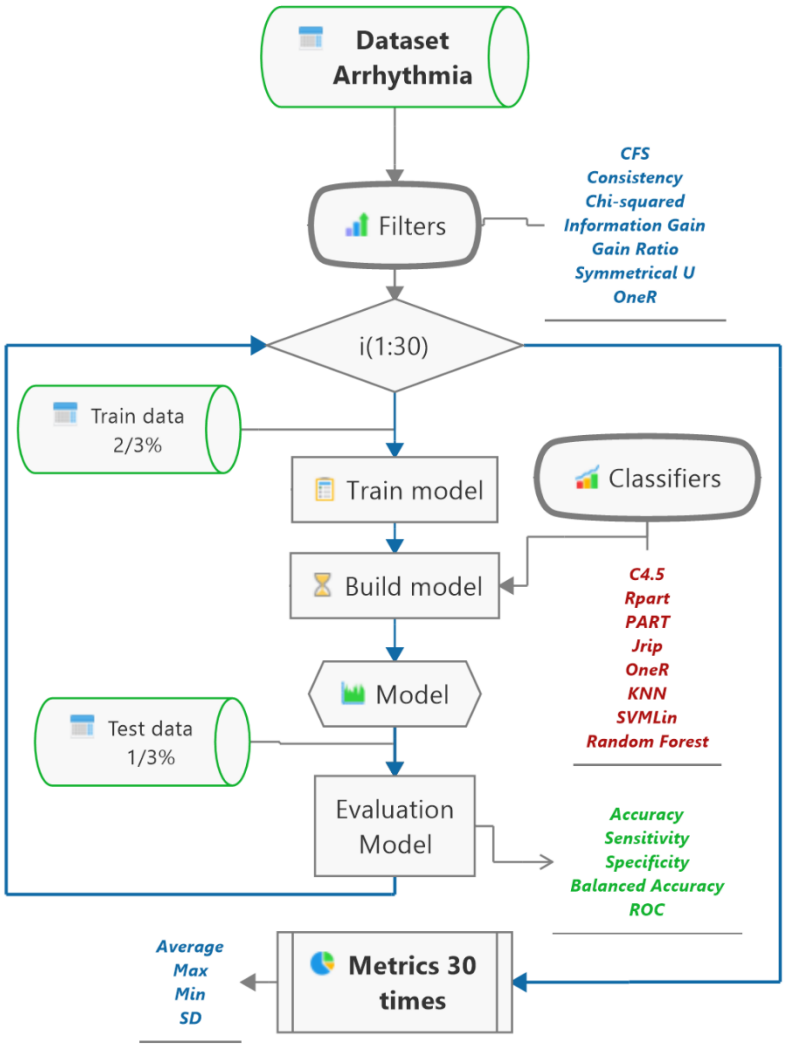

Fig. 1. Proposed methodology for feature selection with filters.

3 Experimental design and results

3.1 Experimental design

As described in section 2.1, the experiments used a cardiac arrhythmia dataset with 278 features and 452 records. Seven filter methods were applied to the dataset. The CFS and Consistency filters produce subsets of the most relevant features. In contrast, the OneR, Chi-Square, Information Gain, Symmetric Uncertainty, and Gain Ratio filters assign scores to each feature, ranking them by relevance. Two new datasets were created for the CFS and Consistency filters, containing the selected features along with the class and all records.

These datasets were split into training (two-thirds) and testing (one-third) sets and evaluated using each classifier over 30 iterations. Each iteration employed a different seed to generate varied training and testing splits. Metrics were computed in every iteration and averaged across all repetitions (Fig. 1).

For the remaining filters, attributes were ranked, and subsets were incrementally selected, starting with the top two, then the top three, and so forth, until all attributes were included. New datasets were generated by sequentially adding the top-ranked attributes, starting with the first two attributes plus the class and all instances, followed by the first three attributes, and so on. These data sets were then divided into training (two-thirds) and test (one-third) sets. Each dataset was applied to the classifiers over 30

iterations, with a different random seed used in each iteration to produce distinct training and testing splits. Metrics were calculated in each iteration and averaged at the end. This procedure was repeated for every filter. The relevant variables with the best performance of the balanced accuracy averaging found with the filter methods were used to implement GA. Once the best subset was identified using the filter methods, the GA function of the GA package (Scrucca, 2013) was used.

This function allows the implementation of an objective function to evaluate each subset within the search space. In addition, genetic operators such as population, number of generations, mutation, and crossover can be defined. For these operators, optimal values were defined by testing different parameters to determine the best performance. They showed the best values with a population of 50 and 100 generations, mutation of 0.1, and crossover of 0.8.

The average of balanced accuracy was the objective function designed to evaluate the subsets generated within the GA function. To obtain this average, each subset was divided into training (2/3) and test (1/3), applied the C4.5 classifier, and repeated 30 times with different seeds to generate different results. For each repetition, the balanced accuracy was calculated by returning the average of each subset (Fig. 2).

In the results of the GA function, each important variable was identified with a value of 1, discarding the least essential variable with a value of 0.

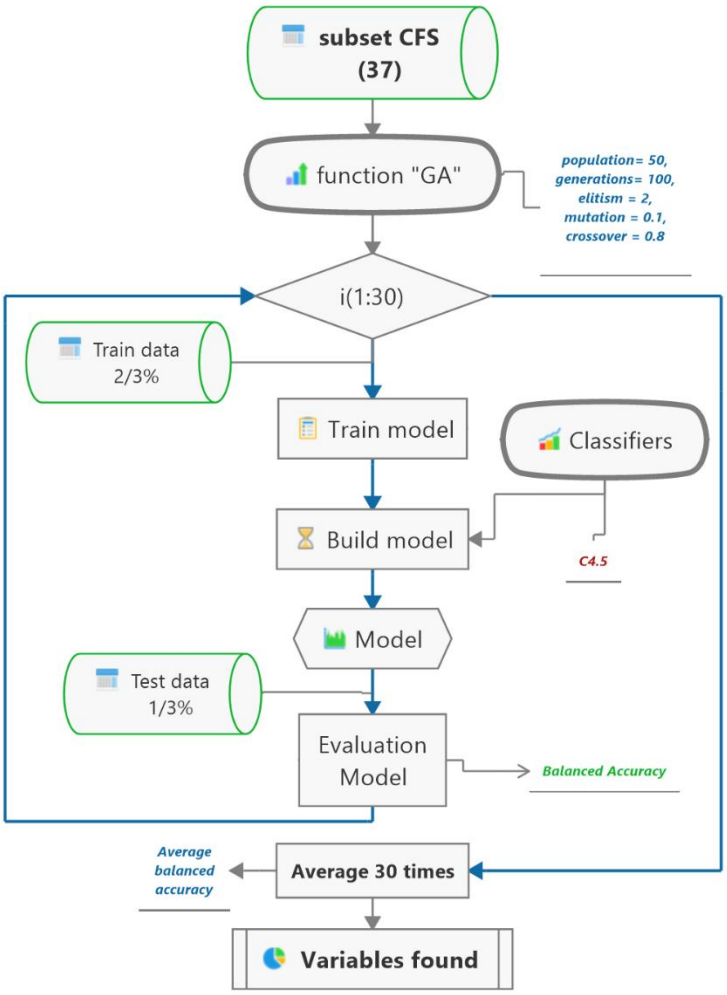

Fig. 2. Objective function to evaluate each subset using the GA package.

After obtaining the subset of the best-selected variables from the GA using the GA function, the classifiers OneR, PART, Rpart, JRip, C4.5, Support Vector Machine, K-Nearest Neighbor, and the combined Random Forest method were implemented. For each classifier, 30 tests were performed using different seeds to obtain the average balanced accuracy (Fig. 3).

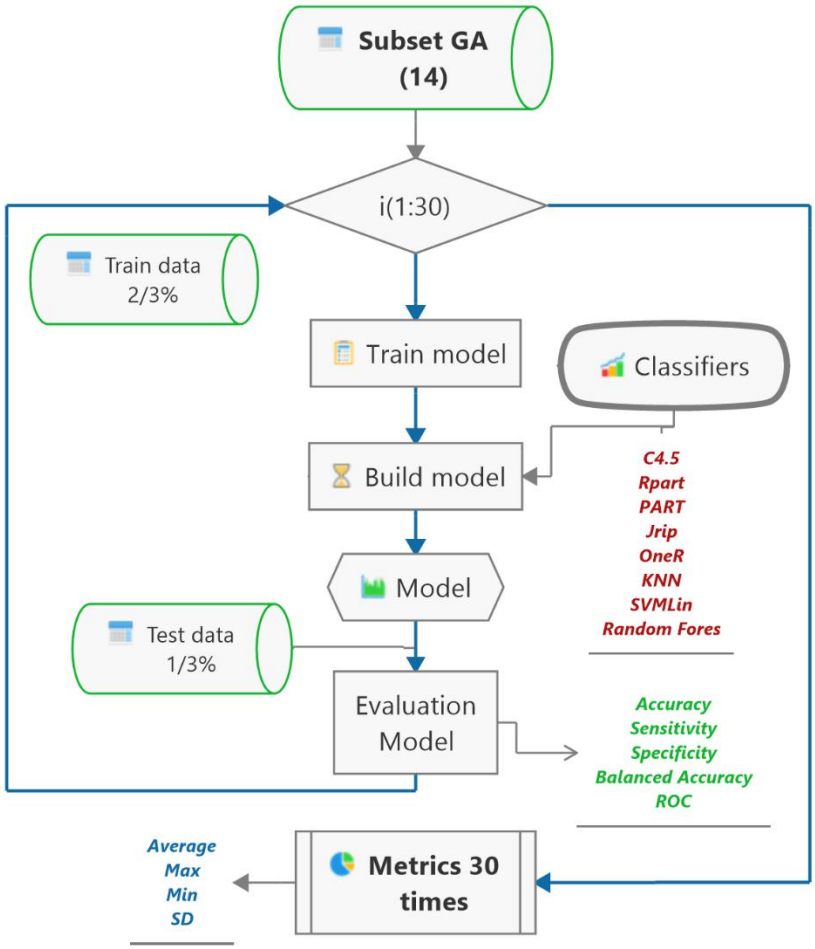

Fig. 3. Feature selection method using GA metaheuristic algorithm.

3.2 Results of filter methods

A subset of relevant variables was obtained using filter methods that increased the average baseline balanced accuracy. Using the random forest classifier, the highest average baseline balanced accuracy using all variables (278) was **0.8147**. The relevant variables selected with the highest average balanced accuracy were those with the CFS filter, with 37 variables and an average balanced accuracy value of **0.8293**. The final values used for the random forest model with CFS were mtry = 11 and ntree = 300.

Table 3 compares the average balanced accuracy for each filter applied with its respective classifier and the total number of relevant variables selected.The performance of the classifiers is observed in terms of the average balanced accuracy. The objective is to identify the best combination of filter methods and classifiers to detect cardiac arrhythmia more effectively, using a reduced number of variables without decreasing model performance.The CFS filter, combined with the Random Forest classifier, achieves the highest average balanced accuracy (0.8293) using a reduced set of 37 variables. This result outperforms the baseline model that uses all 278 variables. Moreover, it fulfills the study's objective by reducing dimensionality and improving model performance in identifying cardiac arrhythmia.The best result in each case is highlighted in bold. The CFS filter, combined with the Random Forest classifier, was selected as the input for the genetic algorithm.

Table 3. Average value using all filters with classifiers.

Filter	Classifiers	OneR	PART	Rpart	JRip	C4.5	SVMLin	KNN	Random Forest
OneR	***Variables found***	2	84	42	32	53	62	20	80
	Accuracy	0.6307	0.7551	0.7627	0.7571	0.7629	0.7702	0.7347	0.8289
	Specificity	0.7593	0.7728	0.7819	0.7844	0.7922	0.9016	0.9551	0.8490
	Sensitivity	0.4797	0.7343	0.7401	0.7251	0.7285	0.6159	0.4758	0.8053
	Balanced accuracy	0.6195	0.7536	0.7610	0.7547	0.7603	0.7588	0.7155	0.8271
	AUC-ROC	0.6319	0.7555	0.7642	0.7583	0.7638	0.7898	0.7926	0.8302
Chi-squared	***Variables found***	2	35	20	40	18	49	17	78
	Accuracy	0.6307	0.7709	0.7696	0.7602	0.7727	0.7729	0.7458	0.8278
	Specificity	0.7593	0.8016	0.7947	0.7860	0.8099	0.9263	0.9383	0.8473
	Sensitivity	0.4797	0.7348	0.7401	0.7300	0.7290	0.5928	0.5198	0.8048
	Balanced accuracy	0.6195	0.7682	0.7674	0.7580	0.7694	0.7595	0.7290	0.8261
	AUC-ROC	0.6319	0.7736	0.7709	0.7608	0.7740	0.8015	0.7889	0.8291
Information Gain	***Variables found***	2	48	21	13	15	62	30	64
	Accuracy	0.6147	0.7729	0.7684	0.7687	0.7702	0.7724	0.7471	0.8309
	Specificity	0.7399	0.8066	0.7922	0.8066	0.8119	0.9086	0.9514	0.8465
	Sensitivity	0.4676	0.7333	0.7406	0.7242	0.7213	0.6126	0.5072	0.8126
	Balanced accuracy	0.6038	0.7700	0.7664	0.7654	0.7666	0.7606	0.7293	0.8295
	AUC-ROC	0.6153	0.7750	0.7697	0.7721	0.7724	0.7944	0.7983	0.8319
Symmetrical Uncertainty	***Variables found***	8	34	37	28	35	62	7	59
	Accuracy	0.5984	0.7722	0.7678	0.7658	0.7738	0.7660	0.7633	0.8302
	Specificity	0.7111	0.8128	0.7885	0.7835	0.8107	0.9070	0.9420	0.8416
	Sensitivity	0.4662	0.7246	0.7435	0.7449	0.7304	0.6005	0.5536	0.8169
	Balanced accuracy	0.5886	0.7687	0.7660	0.7642	0.7706	0.7537	0.7478	0.8292
	AUC-ROC	0.5976	0.7761	0.7685	0.7666	0.7754	0.7891	0.8030	0.8306
Gain Ratio	***Variables found***	2	66	41	48	44	48	42	61
	Accuracy	0.5902	0.7631	0.7804	0.7678	0.7649	0.7678	0.7522	0.8276
	Specificity	0.7267	0.7844	0.8095	0.7959	0.8074	0.9156	0.9576	0.8358
	Sensitivity	0.4300	0.7382	0.7464	0.7348	0.7150	0.5942	0.5111	0.8179
	Balanced accuracy	0.5784	0.7613	0.7779	0.7653	0.7612	0.7549	0.7344	0.8268
	AUC-ROC	0.5888	0.7642	0.7820	0.7694	0.7669	0.7934	0.8057	0.8286
CFS	***Variables found***	37	37	37	37	37	37	37	**37**
	Accuracy	0.5687	0.7744	0.7576	0.7467	0.7744	0.7649	0.7409	0.8311
	Specificity	0.6675	0.8053	0.7786	0.7580	0.7992	0.9247	0.9012	0.8519
	Sensitivity	0.4527	0.7382	0.7329	0.7333	0.7454	0.5773	0.5527	0.8068
	Balanced accuracy	0.5601	0.7718	0.7557	0.7457	0.7723	0.7510	0.7269	**0.8293**
	AUC-ROC	0.5639	0.7769	0.7590	0.7492	0.7756	0.7960	0.7664	0.8323
Consistency	***Variables found***	20	20	20	20	20	20	20	20
	Accuracy	0.5738	0.7589	0.7702	0.7567	0.7598	0.7580	0.7000	0.8267
	Specificity	0.7008	0.7827	0.7909	0.7671	0.7975	0.8638	0.9259	0.8395
	Sensitivity	0.4246	0.7309	0.7459	0.7444	0.7155	0.6338	0.4348	0.8116
	Balanced accuracy	0.5627	0.7568	0.7684	0.7558	0.7565	0.7488	0.6824	0.8250
	AUC-ROC	0.5704	0.7593	0.7713	0.7602	0.7615	0.7690	0.7416	0.8255

Table 4.GA parameters.

Filter	Parameters	Values
GA	Type	Binary
	Population	50
	Generations	100
	Elitism	2
	Cross	0.8
	Mutation	0.1
	Variables found	***12***

Table 5. Average value using GA with classifiers.

Filte	Classifiers	*OneR*	*PART*	*Rpart*	*JRip*	*C4.5*	*SVMLin*	*KNN*	*Random Forest*
GA	***Variables found***	12	12	12	12	12	12	12	**12**
	Accuracy	0.7536	0.8047	0.8164	0.7880	0.8073	0.5938	0.7731	0.8278
	Specificity	0.8988	0.8638	0.8490	0.8263	0.8638	0.7082	0.8963	0.8523
	Sensitivity	0.5831	0.7353	0.7783	0.7430	0.7411	0.4594	0.6285	0.7990
	Balanced accuracy	0.7409	0.7995	0.8136	0.7847	0.8024	0.5838	0.7624	**0.8256**
	AUC-ROC	0.7759	0.8106	0.8194	0.7908	0.8128	0.5920	0.7899	0.8295

Table 6. Relevant variables using GA.

Variable	Description
V15	Heart rate: Number of heart beats per minute
V76	Q wave of channel AVF
V112	Q wave of channel V3
V169	QRSTA of channel DI
V179	QRSTA of channel DII
V197	T wave of the channel AVR
V211	Q wave of channel AVF
V224	R' wave of channel V1
V234	R' wave of channel V2
V250	JJ wave of channel V4
V261	Q wave of channel V5
V277	T wave of channel V6

3.3 Results using the genetic algorithm

The subset found with the GA using the GA function was 12 relevant variables. Table 4 shows the parameters used in the GA function, which aims to optimize the set of relevant variables obtained with the CFS filter. By applying the genetic algorithm metaheuristic, this step aligns with the study's goal of maximizing balanced accuracy and reducing the number of variables required to identify cardiac arrhythmia.

Table 5 shows the average of each classifier, taking the subset chosen with the GA method as input. It is observed that the best average accuracy value is balanced again with the random forest classifier, with an average value of **0.8256**. The Random Forest classifier stands out once again, achieving a balanced accuracy value close to the one obtained with the CFS filter, but this time using a smaller number of variables.Thebestvalueisshown in bold. The final values used for the random forest and GA models were mtry = 2 and ntree = 100.

Table 6 shows the relevant variables selected using the GA metaheuristic and the Random Forest classifier.

4 Conclusions

This study demonstrates the importance of selecting relevant variables using filter methods, genetic algorithms, and classifiers. We applied seven different filters from the FSelector package: OneR, Chi-Square, Information Gain, Symmetric Uncertainty, Gain Ratio, CFS, and Consistency; in addition to the OneR, PART, Rpart, JRip, C4.5, SVMlin, KNN, and Random Forest. With the results obtained, two subsets of relevant variables with better average performance in terms of balanced accuracy were found:

- **Thirty-sevenvariables** were filtered using the CFS filter with a balanced average accuracy of **82.93%** using the combined random forest method.
- **Twelve variables** were selected using the genetic algorithm, with a balanced average accuracy of **82.56%** using the combined random forest method.

When comparing the results, the random forest classifier showed better results. In agreement with studies conducted by other researchers, this work offers a different approach with respect to the number of relevant variables found, demonstrating that selecting relevant variables using filters and genetic algorithms contributes significantly to creating models with optimal performance and a reduced number of variables.

From a medical approach, we aimed to facilitate the diagnosis of cardiac arrhythmia using a reduced set of relevant variables. On the computational side, we seek to promote the use of filter methods using the FSelector package and metaheuristics such as the genetic algorithm. The findings obtained will allow researchers to improve the effectiveness of predictive models.

Future work is expected to complement the results obtained with other metaheuristics that are already in progress, thus allowing a broader comparison of the different metaheuristic algorithms, such as differential evolution, tabu search, and simulated annealing.

References

Al-Saffar, B., Ali, Y. H., Muslim, A. M., & Ali, H. A. (2023). ECG Signal Classification Based on Neural Network. *Lecture Notes in Networks and Systems*, *573 LNNS*, 3–11. https://doi.org/10.1007/978-3-031-20429-6_1

Arias García, S., Hernández Torruco, J., & Hernández Ocaña, B. (2021). Imputación de datos de pacientes con arritmia cardiaca. *Investigación Aplicada, Un Enfoque En La Tecnología*, *6*(2021), 351–359. https://scholar.google.com/citations?view_op=view_citation&hl=es&user=ABDCkcAAAAAJ&cstart=100&pagesize=100&citation_for_view=ABDCkcAAAAAJ:qxL8FJ1GzNcC

Ayar, M., & Sabamoniri, S. (2018). An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm. *Informatics in Medicine Unlocked*, *13*, 167–175. https://doi.org/10.1016/j.imu.2018.06.002

Babatunde, O., Armstrong, L., Leng, J., & Diepeveen, D. (2014). *A genetic algorithm-based feature selection. Edith Cowan University.* https://ro.ecu.edu.au/ecuworkspost2013/653

Betancourt, G. A. (2005). Las máquinas de soporte vectorial (SVMs). (SVMs). *Scientia Et Technica*, *XI*(27), 67–72. https://doi.org/10.22517/23447214.6895

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chakraborty, A., Chatterjee, S., Majumder, K., Shaw, R. N., & Ghosh, A. (2022). *A comparative study of myocardial infarction detection from ECG data using machine learning*. En *Proceedings publicados por Springer* (pp. 257–267).

https://doi.org/10.1007/978-981-16-2164-2_21
Dash, M., & Liu, H. (2000). Feature selection for clustering. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1805*, 110–121. https://doi.org/10.1007/3-540-45571-X_13
Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence*, *151*(1–2), 155–176. https://doi.org/10.1016/S0004-3702(03)00079-1
Elsayad, A. M. (2009). Classification of ECG arrhythmia using learning vector quantization neural networks. *Proceedings - The 2009 International Conference on Computer Engineering and Systems, ICCES'09*, 139–144. https://doi.org/10.1109/ICCES.2009.5383295
Frank, E., & Witten, I. H. (1998). *Generating accurate rule sets without global optimization*. In *Proceedings of the Fifteenth International Conference on Machine Learning* (pp. 144–151).
Guvenir H., A. B. M. H., & Quinlan, R. (1998). *Arrhythmia - UCI Machine Learning Repository*.
Hall, M. A. (1999). *Correlation-based feature selection for machine learning* (Doctoral dissertation). University of Waikato.
Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Elsevier. https://doi.org/10.1016/C2009-0-61819-5
Hechenbichler, K., Schliep, K., & Lizee, A. (2016). *Weighted k-nearest-neighbor techniques and ordinal classification* (SFB 386 Working Paper). Ludwig-Maximilians-Universität München. http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper399.ps
Holte, R. C. (1993). Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. *Machine Learning*, *11*(1), 63–90. https://doi.org/10.1023/A:1022631118932
Hornik, K. (2012). *RWeka odds and ends* (RWeka package vignette). Comprehensive R Archive Network (CRAN). http://ftp.arklinux.org/pub/cran/web/packages/RWeka/vignettes/RWeka.pdf
Jadhav, S. M., Nalbalwar, S. L., & Ghatol, A. (2010). *Artificial neural network based cardiac arrhythmia classification using ECG signal data*. In *Proceedings of the 2010 International Conference on Electronics and Information Engineering (ICEIE)*. https://doi.org/10.1109/ICEIE.2010.5559887
Jangra, M., Dhull, S. K., Singh, K. K., Singh, A., & Cheng, X. (2021). *O-WCNN: An optimized integration of spatial and spectral feature map for arrhythmia classification*. *Complex & Intelligent Systems*. https://doi.org/10.1007/s40747-021-00371-4
Kadam, V. J., Yadav, S. S., & Jadhav, S. M. (2020). *Soft-margin SVM incorporating feature selection using improved elitist GA for arrhythmia classification*. En *Advances in Intelligent Systems and Computing* (Vol. 941, pp. 965–976). Springer. https://doi.org/10.1007/978-3-030-16660-1_94
Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05
Leng, S., Tan, R. S., Chai, K. T. C., Wang, C., Ghista, D., & Zhong, L. (2015). *The electronic stethoscope*. *BioMedical Engineering Online, 14*(1), 1–37. https://doi.org/10.1186/s12938-015-0056-y
Morganroth, J. (1983). Identification of the Patient at High Risk of Sudden Cardiac Death. In *Cardiac Arrhythmias* (pp. 13–19). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-68926-0_3
Nevill-Manning, C. G., Holmes, G., & Witten, I. H. (1995). The development of Holte's 1R classifier. *Proceedings - 1995 2nd New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, ANNES 1995*, 239–242. https://doi.org/10.1109/ANNES.1995.499480
Pandey, S. K., & Janghel, R. R. (2019). ECG Arrhythmia Classification Using Artificial Neural Networks. In *Lecture Notes in Networks and Systems* (Vol. 46, pp. 645–652). Springer. https://doi.org/10.1007/978-981-13-1217-5_63
Parveen, A., Vani, R. M., Hunagund, P. V., & Soher-wardy, M. A. (2021). *Classification of ECG Arrhythmia Using Different Machine Learning Approach* (pp. 319–325). Springer, Singapore. https://doi.org/10.1007/978-981-33-4604-8_25
Roopa, C. K., Harish, B. S., & Aruna Kumar, S. V. (2018). *A novel method of clustering ECG arrhythmia data using robust spatial kernel fuzzy C-means*. *Procedia Computer Science, 143*, 133–140. https://doi.org/10.1016/j.procs.2018.10.361
Salzberg, S. L. (1994). *C4.5: Programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993* [Book review]. *Machine Learning, 16*(3), 235–240. https://doi.org/10.1007/BF00993309
Scrucca, L. (2013). *GA: A package for genetic algorithms in R*. *Journal of Statistical Software, 53*(4), 1–37. https://doi.org/10.18637/JSS.V053.I04
Therneau, T., Atkinson, B., & Ripley, B. (2022). *rpart: Recursive partitioning and regression trees* (R package). Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=rpart
Xu, S. S., Mak, M. W., & Cheung, C. C. (2017). Deep neural networks versus support vector machines for ECG arrhythmia classification. *2017 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2017*, 127–132. https://doi.org/10.1109/ICMEW.2017.8026250
Zheng, Z., Wu, X., & Srihari, R. (2004). Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explorations Newsletter*, *6*(1), 80–89. https://doi.org/10.1145/1007730.1007741