



Early detection of Ventricular Bigeminy/Trigeminy rhythms

Bruno Rodrigues de Oliveira ^{1*}, Marco Aparecido Queiroz Duarte ², Jozue Vieira Filho ³

¹Pantanal Editora, Rua Abaete, 83, Sala B, Centro, Nova Xavantina-MT. 78.690-000, Brasil.

²Universidade Estadual de Mato Grosso do Sul (UEMS), Unidade de Cassilândia, Rodovia MS 306 - Km 6,4, Cassilândia-MS, 79540-000, Brasil.

³Engenharia de Telecomunicações e Aeronáutica, Universidade Estadual Paulista (UNESP), Av. Profa. Isette Correa Fontão, 505 - Jardim das Flores, São João da Boa Vista - SP, 13876-750, Brasil.

*Corresponding author. E-mail: bruno@editorapantanal.com.br

ABSTRACT. Premature Ventricular Contraction is an arrhythmia that can be associated with several cardiac disorders that affect from 40% to 75% of the general population. Premature Ventricular Contraction occurrence is diagnosed from the Electrocardiogram. If in an Electrocardiogram one (or two) Premature Ventricular Contraction occurs between two Normal heartbeats, then there is a Ventricular Bigeminy (or Trigeminy). The prevalence of Ventricular Bigeminy/Trigeminy rhythms is associated with angina, hypertension, congestive heart failure and myocardial infarction. In this work it is proposed a new approach for these rhythms early diagnosis using Decision Tree models. The proposed approach uses the information before occurrence of Ventricular Bigeminy/Trigeminy, i.e., the number of normal and abnormal heartbeats and the heart rhythm. In order to rhythm prediction, the models obtained from Random Forest algorithm, induced by cross-validation approach, are used. Proposed approach predicted Ventricular Bigeminy/Trigeminy occurrence with accuracy, sensitivity and specificity of 98.94%, 96.28% and 99.83, respectively. Furthermore, the results showed that the Ventricular Bigeminy/Trigeminy is preceded for Normal, Atrioventricular Junctional and Paced heart rhythms in most of the examples. Besides that, it is presented a simple algorithm for decision about the occurrence of Ventricular Bigeminy/Trigeminy rhythms.

Keywords: Premature Ventricular Contraction; Random Forest; Machine Learning.

DOI: <https://doi.org/10.33837/msj.v5i1.1525>

Received: November 18, 2021. Accepted: February 7, 2022.

Associate editor: Anderson Rodrigo da Silva

INTRODUCTION

Heart rhythm depends on the specific initialization and propagation of the electrical impulses from the specialized cardiac cells. Normally, these impulses arise from sinus node (Natural pacemaker). But, for a variety of reasons, such as valve disorders, high blood pressure, cardiomyopathy, injury from a heart attack, use of drugs, among others, they can arise in another site in the heart (Gerry & Lemery, 2018). In these cases, the observed effect is that the heart beats very fast, very slowly or irregularly. When the ventricles assume the function of pacemaker, Premature Ventricular Contraction (PVC) is characterized. This arrhythmia is due to three main effects: abnormal impulse formation, reentry and triggered activity (Latchamsetty & Bogun,

2015). It has been studied by several researchers, who have proposed algorithms for its automatic recognition based on two template-matching procedures using the correlation coefficients (Li et. al, 2014); eigenvectors obtained from Principal Component Analysis and Linear regression analysis (Hadia et. al, 2017); Gaussian process classifiers, wavelet and S transforms (Bazi et. al, 2013); geometrical features and ensemble of machine learning using analytic hierarchy process (Oliveira et. al, 2019). Such studies are very important because PVC can be associated to the risk of sudden death (Fred, 2009) and a high frequency of its occurrence can lead to hemodynamic problems (Garcia & Miller, 2004).

PVC occurrence is more frequent in people with higher risk of coronary heart disease (Latchamsetty & Bogun, 2015). It is associated to structural heart diseases and represents an increased risk of cardiovascular diseases and sudden death (Amir et. al, 2020, Ahn, 2013). Although PVC occurrence can be considered benign in people without any structural heart disease, some studies suggest that PVC can induce cardiomyopathy even in healthy subjects (Ahn, 2013). This arrhythmia was

Copyright © The Author(s).

This is an open-access paper published by the Instituto Federal Goiano, Urutai - GO, Brazil. All rights reserved. It is distributed under the terms of the Creative Commons Attribution 4.0 International License.



detected in 40% to 75% of general population evaluated by a 24-hour or 48-hour Holter Monitoring (Ahn, 2013).

For measuring the electrical activity of the heart, it is commonly used the Electrocardiograph that provides the Electrocardiogram (ECG), which is composed by the waves P (atrial depolarization), QRS complex (ventricular depolarization) and T (ventricular repolarization). If an alternating heartbeat pattern composed of Normal and PVC is observed in an ECG segment, then there is a Ventricular Bigeminy rhythm when there exists one PVC between two normal heartbeats. Or a Ventricular Trigeminy rhythm, for two PVC between two normal heartbeats (Chugh, 2012). The presence of repetitive Ventricular Bigeminy/Trigeminy was associated with certain clinical characteristics as angina, hypertension, congestive heart failure and myocardial infarction (Weaver, Cobb & Hallstrom, 1982). In Wang et. al (2021) the prevalence of Bigeminy/Trigeminy was significantly higher in patients with obstructive sleep apnea.

Some researchers have proposed methodologies for cardiac rhythm recognition. Andersen, Peimankar & Puthusserypady (2019) proposed a real-time methodology combining the Convolutional- and Recurrent-Neural Networks using RR intervals to classify ECG segments as Atrial Fibrillation (AF) or Normal sinus rhythm. Their methodology extracts high level features from segments of RR intervals directly from the data. The results, over three datasets including a total of 89 subjects, achieved sensitivity and specificity of 98.98% and 96.95%, respectively.

Teplitzky, McRoberts & Ghanbari (2020) presented the BeatLogic® platform designed for heartbeat and rhythm detection/classification, employing Deep Learning models such as convolutional networks. This platform is divided in two main networks: RhythmNet and BeatNet. The first network detects/classifies rhythms as Sinus rhythm, Atrial fibrillation/flutter, Supraventricular tachycardia, Junctional rhythm, Second-degree heart block type 1, Second-degree heart block type 2, Third-degree heart block, and other. The second network detects heartbeats, artifacts and pause annotation, and classifies the heartbeats. In the detection stage the methodology achieves sensitivity and positive predictive values of 99.84% and 99.78%, respectively. In the classification stage, the performance of recognizing the Ventricular ectopic beats (or PVC) was 89.4% and 97.8% of sensitivity and positive predictive, respectively.

In a similar way, Goodfellow et al. (2018) implemented a 13-layer convolutional network for classification among Normal, Atrial Fibrillation (AF), and other rhythms in single lead ECG. The network includes dilated convolutions, max pooling, ReLU

activation, batch normalization, and dropout. The average results were 84%, 85%, 88% for precision, recall and accuracy, respectively. The worst classification occurred among the other/AF classes in 15.86% of the instances.

Unlike previous works, Hajeb-Mohammadipour et. al (2018) developed an automated approach to identify many types of arrhythmic episodes using a classical Machine Learning (ML) algorithm, namely, binary Decision Tree, support vector machine and many other types of signal characterization, such as temporal and frequency representation, non-linear analysis and image-based phase plot. This approach is divided in four stages: (1) classification between Ventricular Fibrillation (VF) and non-VF, which includes Ventricular Bigeminy/Trigeminy; (2) classification between AF and non-AF; (3) PVC detection on every non-AF; (4) determination of the PVC patterns, to categorize the types of sinus arrhythmias and normal sinus rhythm. The obtained results in a balanced dataset were 95.1%, 94.5% and 94.2% of accuracy, sensitivity and specificity, respectively.

A common attribute of the aforementioned works is that such methodologies analyze an ECG segment and then decide on its rhythm. In other words, these algorithms only predict the past rhythm, because the ECG is a time series. However, early detection is very important for: (1) remote monitoring systems (RMS) (Hernández-Madrid et al, 2014), since according to Ricci et. al (2013), RMS prevents multiple face-to-face visits to the hospital and enables early detection of problems and continuous follow-up, improving safety and quality of life; (2) heart failure (HF) diagnosis (Tsai et al, 2020), because HF remains a major cause of mortality and also leads to enormous social problems around the world (Neubauer, 2007); and (3) development of applications for implantable cardiac defibrillators (Taye et al, 2019).

Other works have focused in early rhythm prediction. Lee et. al (2016) disclose a method for prediction of ventricular tachycardia one hour before its occurrence by means of artificial neural networks (ANN) (with backpropagation learning rule and a perceptron structure), using until 14 parameters obtained from heart rate variability (HRV) and respiratory rate variability. The performance achieved sensitivity of 88%, specificity of 82%, and AUC of 93%.

Taye et. al. (2019) presented a method for early ventricular tachyarrhythmia prediction using HRV from 120 seconds and QRS complex shape features. This method detects the ventricular fibrillation 30 seconds before its occurrence using two ANN trained with HRV and QRS features. Employing 10-fold cross-validation, the results were 72% of accuracy using 11 HRV features, while using four QRS complex shape features the accuracy was 98.6%.

Cappiello et. al (2015) proposed a novel statistical index for the early diagnosis of ventricular arrhythmia using the time delay phase-space reconstruction technique from ECG, predicting an impending arrhythmia from 356 ECG heartbeats, on average, before its onset. The statistics calculated in phase space diagrams were: mean, standard deviation, coefficient of variation, skewness and kurtosis. The achieved results were 96.88% of sensitivity, 100% of specificity and 98.44% of accuracy.

Since PVC episodes and, consequently, Ventricular Bigeminy/Trigeminy rhythms are associated with severe cardiac disorders, it is useful that rhythm recognition methods are able to predict also these rhythms specifically. The methods mentioned in last paragraphs do not focus on the early recognition of Ventricular Bigeminy/Trigeminy rhythms. Only Hajeb-Mohammadalipour et. al (2018) investigated these rhythms, but only in opposition to the AF class, that is, the focus is not on Ventricular Bigeminy/Trigeminy.

For these reasons, in this work a new approach is proposed to predict the Ventricular Bigeminy/Trigeminy rhythms using only information from the ECG segment before occurrence of Ventricular Bigeminy/Trigeminy. For this purpose, it is used the Random Forest algorithm to generate Decision Tree models. The mentioned information is composed only of the previous rhythm and the amount of Normal and Abnormal heartbeats, making it a very simple and low-cost method.

MATERIALS AND METHODS

Training data

Two data bases are used in this research: MIT-BIH Arrhythmia (Moody & Mark, 2001) (MITDB) and Long-Term Atrial Fibrillation (Petrutiu, Sahakian & Swiryn, 2007) (LTAfDB), both provided by the PhysioNet repository (Goldberger et al, 2000). MITDB contains 48 half-hour excerpts from two-channel ambulatory ECG recordings, digitized at 360 samples per second per channel, with 11-bit resolution over a 10 microvolt (mV) range, obtained from 47 subjects. LTAfDB includes 84 long-term ECG recordings digitized at 128 Hz with 12-bit resolution over a 20mV range, with durations normally varying from 24 to 25 hours. Both databases contain information about the heartbeat type and localization (cardiac cycle) and the rhythm, which are described in Table 1.

For the purpose of this research, the heartbeats different from N (Normal beat in Table 1) are accounted as Abnormal. Since not all ECG recordings have Ventricular Bigeminy/Trigeminy occurrences, only the records shown in Table 2 were selected.

Table 1. Type of heartbeats and rhythms present in MITDB and LTAfDB databases

Heartbeat type		Rhythm	
Code	Description	Code	Description
N	Normal beat	rAB	Atrial Bigeminy
L	Left bundle branch block beat	rAFIB	Atrial fibrillation
R	Right bundle branch block beat	rAFL	Atrial flutter
B	Bundle branch block beat	rB	Ventricular Bigeminy
A	Atrial premature beat	rBII	2° heart block
a	Aberrated atrial premature beat	rIVR	Idioventricular
J	Nodal premature beat	rN	Normal sinus
S	Supraventricular premature	rNOD	Nodal (A-V junctional)
V	Premature ventricular contraction	rP	Paced
r	R-on-T premature ventricular contraction	rPREX	Pre-excitation
F	Fusion of ventricular and normal beat	rSBR	Sinus bradycardia
e	Atrial escape beat	rSVTA	Supraventricular tachyarrhythmia
j	Nodal (junctional) escape beat	rT	Ventricular Trigeminy
n	Supraventricular escape beat	rVFL	Ventricular flutter
E	Ventricular escape beat	rVT	Ventricular tachycardia
/	Paced beat		
f	Fusion of paced and normal beat		

Table 2. ECG recordings select for induced/test the prediction models.

Database	ECG recordings	Number of rhythms		
		rB	rT	rN
LTAfDB	01, 07, 08, 13, 16, 22, 23, 28, 45, 47, 51, 58, 72, 100, 101, 103, 119, 121, 122, 200	2638	733	13267
MITDB	106, 119, 124, 200, 201, 203, 207, 208, 210, 213, 214, 217, 219, 221, 223, 228, 233	221	83	310

Machine learning framework

In order to develop a model to predict rhythms Machine Learning (ML) algorithms are employed. These algorithms map the input space, whose vectors represent characteristics of the cardiac cycle, to an output space with scalar elements, where each element describes a particular rhythm. In the context of this work, let $T = \{(\mathbf{x}_k, y_k)\}_{k=1}^K$ be a dataset with K instances, where $\mathbf{x}_k = (x_{1k}, x_{2k}, x_{3k}) \in \mathbb{R}^3$ and $y_k \in \{0, 1\}$ are the feature vector (input) and the class (output), respectively. Output values are codes for the

rhythms, i.e., Normal rhythm is encoded as 0 and Ventricular Bigeminy/Trigeminy as 1. About the input, the third component, x_{3k} , also encodes the rhythm, but only for previous ECG segment. In this case, all rhythms in Table 1 are considered. On the other hand, the first two components, x_{1k} and x_{2k} are real numbers varying in the range $[0, 1]$. These values are related to the number of Normal and Abnormal heartbeats, divided by the total of heartbeats.

In this research the well-known Random Forest (RF) algorithm is employed. This algorithm is an ensemble approach that combines different Decision Tree (DT) models considering randomly chosen features subspaces. Due to randomness, the split points in the DT are different for each initialized model. The final decision on the prediction is obtained by merging the prediction from the individual DT through a majority vote scheme (Awad & Khanna, 2015, Breiman, 2001). RF overcomes individual Decision Trees because overfitting is not a problem (Breiman, 2001) and also improves the accuracy (Ho, 1998). Here, this algorithm is chosen because it generates models whose structure is easy to interpret, considering a single DT model. It is important for the validation of models by physicians.

In order to evaluate the rhythm recognition task, classical metrics are used: accuracy, specificity and sensitivity, given by $Acc = \frac{TP+TN}{TP+TN+FP+FN}$, $Sp = \frac{TN}{TN+FP}$ and $Se = \frac{TP}{TP+FN}$, respectively, where TP , TN , FP and FN are true positive, true negative, false positive and false negative, respectively. Sensitivity is also called as True Positive Rate (TPR), whereas False Positive Rate (FPR) is given by $FPR = \frac{FP}{FP+TN}$.

These rates are also used in Receiver Operating Characteristics (ROC) analysis, where each ordered pair (FPR, TPR) is a point in a two-dimensional ROC space that represents the recognition performance. In order to generate ROC curves a score of the recognition model is employed as a threshold. For each threshold, different points are created, forming the ROC curve. Generally, this threshold is associated to prediction probability. The Area Under Curve (AUC) can be associated to each ROC curve (Tom, 2006). AUC equal to 1 means the best performance, since all positive instances are predicted correctly and no negative instance is mistaken as a positive one. Commonly, in the ROC graphs there is a diagonal line, where the pair of components (FPR, TPR) has the same value. A model that produces points under this line is considered as one that cannot use the information learned. If the points are over this line, then this model have no information about the class (Tom, 2006).

Proposed approach

The proposed approach¹ requires some information already available in the databases, such as: QRS (or R wave) position, heartbeat classification and rhythm change position. Therefore, the first procedure is to create a dataset from MITDB and LTAfDB databases separating inputs and outputs according to the ECG annotations. Therefore, rhythm change markings are used, represented by symbol “+” in these databases. Afterwards, Ω_Q heartbeats are taken before each Q rhythm change mark and calculated the number of Normal (N_{Ω_Q}) and Abnormal (A_{Ω_Q}) heartbeats, normalized as $N'_{\Omega_Q} = N_{\Omega_Q}/\Omega_Q$ and $A'_{\Omega_Q} = A_{\Omega_Q}/\Omega_Q$. It is done independently for each ECG recording. Ω_Q parameter is important because it is related to the length of the ECG segment and the larger it is, the more a specific type of heartbeat can occur than other. Therefore, the aforementioned normalization procedure is required.

At end, the k -th input is $x_k = (N'_k, A'_{\Omega_k}, \alpha_k)$ and the output is $y_k = \beta_k$, where α_k and β_k are the rhythms from the previous and predicted ECG segments, respectively, where $k = 1, \dots, K$, such that K is the total of rhythm changes. α_k may have any rhythms from Table 1, column code. On the other hand, β_k takes only the values 0 and 1 which encode rN and rB (or rT), respectively. Figure 1 illustrates the proposed approach.

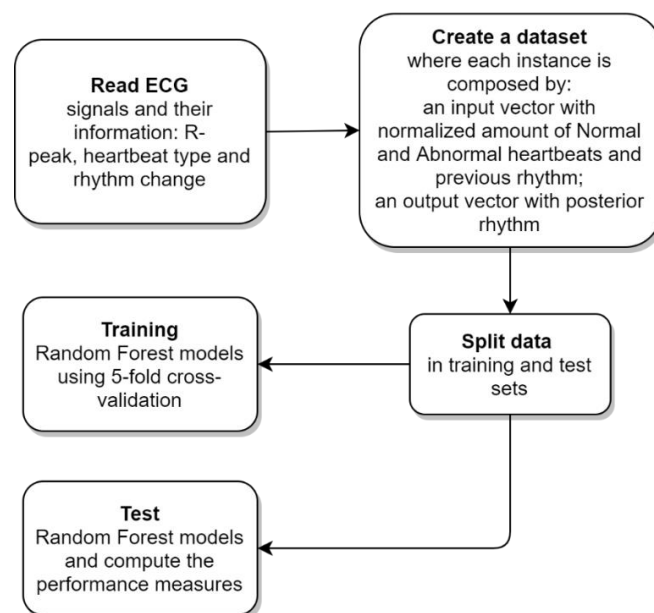


Figure 1. Block diagram of the proposed approach.

¹ Python source codes for the proposed approach are available at: <https://github.com/brunobro/early-detection-of-ventricular-bigeminy-trigeminy-rhythms>

Since the rhythm's values are categorical, it is implemented a function to encode them into numeric values. Furthermore, if the previous ECG segment contains more than one rhythm, then it is discarded. Therefore, the higher Ω_Q parameter is, the more segments are discarded, and consequently there will be less instances to learn the models.

In order to learn and validate the RF models, the 5-fold cross-validation is employed, which consists in subdividing T dataset into disjoint sets called induction and test, five times. Since each fold generates a specific model with its own performance values, overall performance is given by the average.

Lastly, for the computational implementation, three popular and efficient libraries written in Python language were used: Numpy (Harris, 2020), Pandas (McKinney, 2010) and Scikit-learn (Pedregosa et al, 2011). In order to fit the DT models induced by RF algorithm, the default parameters set in Scikit-learn are used. The most relevant is the number of estimators which is related to the number of DT models considered in the ensemble. Here, it is set as 100.

RESULTS

Patterns found in the datasets are shown in Table 4, where the occurrences of heartbeats are described. It shown the variability of previous and posterior rhythms as well as the heartbeats type. It reveals that it is difficult to obtain a model for mapping input to output. For example, first row of Table 3 reveals that there are 11 occurrences of the pattern with input heartbeat N (Normal beat) and rhythm rAB (Atrial Bigeminy) and output rB (Bigeminy). On the other hand, there are also 11 occurrences of a similar pattern, different only in relation to heartbeat which is of the type A (Atrial premature beat). Table 3 also shows that the most frequent pattern consists of the rN rhythm

preceded by Normal heartbeat and rAFIB rhythm (4th row).

Since the choice of Ω_Q (number of previous heartbeats) changes the number of instances used to induce the models, in Table 5 it is shown the number of rhythms when Ω_Q is fixed at 5, 10 or 15 heartbeats. These values correspond to 4, 8 and 15 seconds, respectively, taking into account an average duration of 0.8 seconds for the cardiac cycle. In addition, this table also shows the average percentage of occurrences of Normal and Abnormal heartbeats used in the input, e.g., for $\Omega_Q = 5$, on average 30% of previous heartbeats are Normal and 70% are Abnormal. Generally, there exist much more Normal than Abnormal heartbeats, but in the present study, it must be remembered that only ECG recordings with Ventricular Bigeminy and Trigeminy rhythms (see Table 2) are employed. Therefore, the percentage of Normal heartbeats is lower.

Furthermore, the values presented in Table 4 clarify that there are much more Normal rhythm occurrences than Ventricular Bigeminy and Trigeminy in the previous ECG segment. On the other hand, the frequency of Normal heartbeats is less than Abnormal, for ECG segments that precede changes in rhythm, as it can be seen in Table 3. Due to this characteristic and due to the disposal of ECG segments, mentioned in previous section, there is more variation (when Ω_Q changes) in the number of previous rhythms than posterior rhythms, considering rB and rT. Furthermore, it is important to note that the same rhythm can be computed for previous and posterior segments, since ECG is a time series.

Table 4. Input and output occurrences found in the ECG recordings.

Row	Previous ECG segment (input)			Posterior ECG segment (output)	
	Heartbeat occurrence			Rhythm	Rhythm
	N	V	A		
1	11	0	11	rAB	rB
2	14546	0	18053	rAB	rN
3	62363	1298	2	rAFIB	rB
4	307299	5276	11	rAFIB	rN
5	1967	190	0	rAFIB	rT
6	9176	11685	0	rB	rN
7	11	11	0	rB	rT
8	234774	16774	7841	rN	rB
9	102759	6684	2365	rN	rT
10	88	0	2	rSBR	rB
11	36854	273	256	rSBR	rN
12	45	0	0	rSBR	rT
13	0	0	26	rSVTA	rB
14	0	0	12228	rSVTA	rN
15	0	0	6	rSVTA	rT
16	5	3	0	rT	rB
17	4238	2839	0	rT	rN
18	0	78	0	rVT	rN
Total	774136	45111	40801		

Table 5. Number of rhythms and proportion of heartbeats type.

Code rhythm	Previous heartbeats (Ω_0)		
	5	10	15
Occurrences in the posterior segments (output)			
rN	9,520	2,687	1,563
rB	2,437	1,995	1,703
rT	761	639	566
Total	12,718	5,321	3,832
Occurrences in the previous segments (input)			
rAB	3,591	721	414
rAFIB	702	496	440
rB	2,731	516	273
rIVR	3	1	1
rN	3,093	2,560	2,204
rNOD	2	0	0
rP	8	7	7
rSBR	1,391	578	344
rSVTA	390	75	45
rT	797	360	98
rVFL	5	5	4
rVT	5	2	2
Total	12,713	5,319	3,830
Heartbeat's type		Average percentage occurrences	
Normal	30%	37%	38%
Abnormal	70%	63%	62%

Table 6. Results for 5-fold cross-validation in the training dataset.

Previous heartbeats (Ω_0)	Fold	Accuracy	Sensitivity	Specificity
5	1	0.9914	0.9668	0.9996
	2	0.9919	0.9687	0.9997
	3	0.9919	0.9691	0.9996
	4	0.9913	0.9656	0.9999
	5	0.9980	0.9926	0.9999
	Avg.*		0.9929	0.9726
10	1	0.9871	0.9820	0.9921
	2	0.9864	0.9739	0.9986
	3	0.9875	0.9829	0.9921
	4	0.9864	0.9820	0.9907
	5	0.9993	0.9986	1.0000
	Avg.*		0.9893	0.9839
15	1	0.9843	0.9758	0.9968
	2	0.9843	0.9763	0.9960
	3	0.9856	0.9763	0.9992
	4	0.9843	0.9758	0.9968
	5	0.9993	0.9989	1.0000
	Avg.*		0.9876	0.9806

*Average of the five values of performance

Table 7. Results for 5-fold cross-validation in the test dataset.

Previous heartbeats (Ω_0)	Fold	Accuracy	Sensitivity	Specificity
5	1	0.9988	0.9953	1.0000
	2	0.9965	0.9875	0.9995
	3	0.9965	0.9875	0.9995
	4	0.9976	0.9969	0.9979
	5	0.9575	0.8466	0.9947
	Avg.*		0.9894	0.9628
10	1	0.9934	1.0000	0.9870
	2	0.9972	0.9981	0.9963
	3	0.9915	0.9962	0.9870
	4	0.9962	1.0000	0.9926
	5	0.9173	0.8406	0.9926
	Avg.*		0.9791	0.9670
15	1	0.9987	1.0000	0.9968
	2	0.9987	0.9978	1.0000
	3	0.9935	0.9978	0.9872
	4	0.9974	1.0000	0.9936
	5	0.8995	0.8326	0.9968
	Avg.*		0.9775	0.9656

To find out how much the performance changes when Ω_Q increases, it is evaluated in the range from 10 to 150. Figure 2 reveals that all performance measures decrease when Ω_Q increases, but this reduction is not constant. Moreover, Sensitivity suffers less oscillation than other performance measures.

As mentioned in Section Machine learning framework, RF algorithm is an ensemble of DT models. Thus, the class is predicted taking into account the contribution of each DT model. Performance measures shown in Tables 6 and 7 are the outputs of the ensemble. In order to verify each contribution, the performance measures are computed for each one of the 100 trees and the results are shown in the boxplot, Figure 3. For this experiment 70% and 30% of all instances are considered for induction and test, respectively.

Performance measures presented in Tables 6 and 7 are sensitive to the balance of rhythms classes. On the other hand, ROC curves, that associate TPR and FPR, are insensitive to changes in class distribution (Tom, 2006).

Figure 4 shows ROC curves and the related AUC for each fold with $\Omega_Q = 5$, mean ROC and the standard deviation (σ) of the five folds.

One of the characteristics of the DT algorithms is their ability in measuring the features importance, i.e., they return the weight that a feature has at predicting a class (output) based on the criterion used to split the tree. Table 8 describes the percentage importance that each feature has, according to the number of previous heartbeats, using 30% of the data for test and 70% for induction, without implementing cross-validation method.

From Tables 6 and 7, it is clear that the induced models make prediction errors for some instances. In order to diagnose the most common prediction errors, in Table 9 the frequencies of errors are shown, comparing the rhythm predicted with the true rhythm. Analogously to previous results, these frequencies are described for different Ω_Q values. It is noted that the major errors happen when previous rhythms are equal to rAFIB and true rhythm is equal to rB or rT.

One of the advantages of DT models is their association with a tree structure, which is useful for human interpretation, unlike other ML methods, such as neural networks. In order to obtain a simple tree structure, in the DT algorithm the max depth parameter is set to 2. Thus, the tree is expanded until it obtains two branches. Employing DT approach with 70% for induction and 30% for test, the tree structure is shown in Figure 5. It can be converted in an algorithmic language, as presented in Algorithm 1, which results in accuracy, sensitivity and specificity close to 99.19%, 99.93% and 96.29%, respectively, in the test dataset.

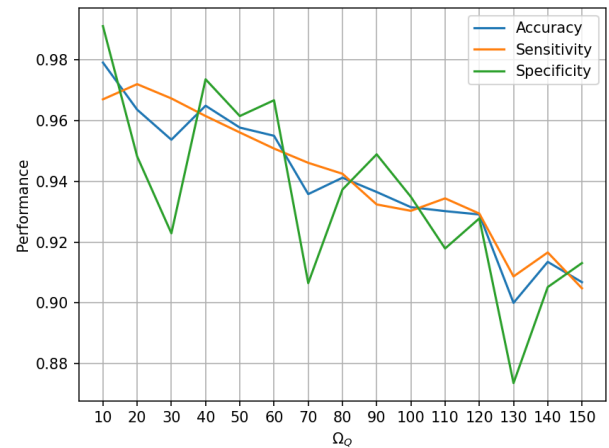


Figure 2. Performances measures when Ω_Q increases.

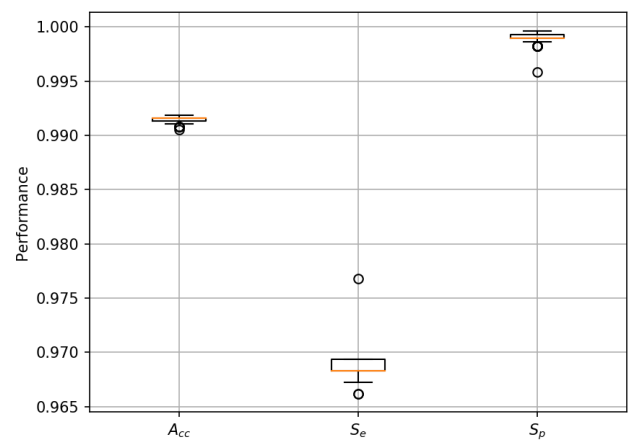


Figure 3. Boxplot of the performance measures for 100 DT models, generated in RF.

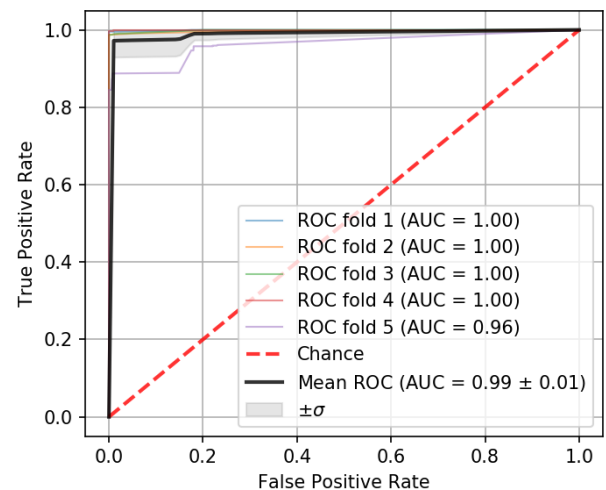


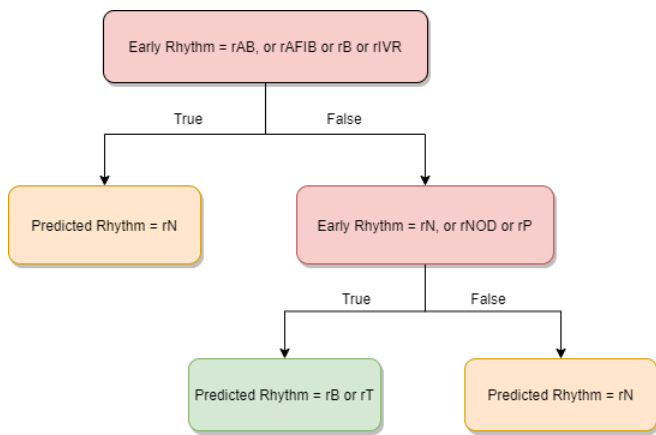
Figure 4. ROC curves and AUC for each model for 5-fold cross-validation.

Table 8. Feature importance given by RF algorithm.

Previous heartbeats (Ω_Q)	Features (input) importance		
	N'_{Ω_Q}	A'_{Ω_Q}	α_k
5	15%	19%	66%
10	16%	22%	62%
15	17%	22%	61%

Table 9. Prediction errors for different number of previous heartbeats.

Previous heartbeats (Ω_Q)	Previous rhythm	Truerhythm	Predicted rhythm	Occurrences
5	rAFIB	rB or rT	rN	27
	rB	rN	rB or rT	1
	rNOD	rN	rB or rT	2
	rSBR	rB or rT	rN	2
10	rAFIB	rN	rB or rT	2
	rB	rB or rT	rN	16
15	rAB	rB or rT	rN	1
	rAFIB	rN	rB or rT	1
	rAFIB	rB or rT	rN	11
	rT	rB or rT	rN	1

**Figure 5.** A tree structure model generated using 70% of the data for training and 30% for testing.

R_e : early rhythm
 R_p : predicted rhythm

if $R_e \in \{rAB, rAFIB, rB, rIVR\}$ then
 $R_p = rN$

else
 if $R_e \in \{rN, rNOD, rP\}$ then
 $R_p = rB$ or rT

 else
 $R_p = rN$

Algorithm 1. Algorithm obtained from a DT model.

DISCUSSION

The results presented in Tables 6 and 7 show a high performance regardless Ω_Q parameter and dataset. Therefore, for fold number 5, decreases in Accuracy and Sensitivity were observed, in relation to test dataset. Nevertheless, this decrease was less than 0.05. These results mean that the models learned by RF algorithm using the proposed features have very low generalization errors. Moreover, since for $\Omega_Q = 5$ greater performance was achieved, few information,

related to ECG segment, is necessary for Bigeminy/Trigeminy recognition.

On the other hand, curves in Figure 2 reveal that the performances decrease when Ω_Q increases too much. This is due to the reduction in the number of instances used to learn the models, as already explained in Section Machine learning framework (Table 5). Also, it is important to remark that the Specificity metric is the only one with greater variation. Such results suggest that the higher is Ω_Q , the worse are the models for Bigeminy/Trigeminy classification.

The analysis of the feature's importance (Table 8) reveals that the previous rhythm has greater importance in the prediction models than the number of Normal and Abnormal heartbeats. The sum of the percentual importance of Normal (N'_{Ω_Q}) and Abnormal (A'_{Ω_Q}) features is inferior to the importance of the previous rhythm (α_k), for all Ω_Q . It indicates that, for the used dataset and DT models learned, the previous rhythm is decisive for the posterior rhythm prediction. In addition, we see that Algorithm 1 (which has an accuracy of 99.19% and represents a simple DT model) does not use information about the number of heartbeats.

Tree structure in Figure 5 and Algorithm 1 also shows that the Ventricular Bigeminy/Trigeminy is only preceded by the rhythms rN, rNOD or rP. This decision is made with a precision (Specificity) of 96.29%. However, this Specificity decreases when compared to results obtained using 100 DT models in RF, due to the lack of information about the number of Normal and Abnormal heartbeats (see Figure 2). Therefore, rB and rT rhythms are more accurately predicted when the number of heartbeats is known. In general, Algorithm 1 explains that when there is a change of pace to Ventricular Bigeminy/Trigeminy, the previous rhythm must be rN, rNOD or rP. On the other hand, if the change of rhythm is for rN, then the previous rhythm can be rAB, rAFIB, rB or rIVR, but if it is none of these, then it must be different than rN, rNOD or rP. It is interesting to note that rN and rB rhythms can occur interspersed. In the datasets used in this work the occurrences of rN before rB and rB before rN were 2353 and 2728 times, respectively, which together correspond to 40% of total instances.

Besides that, Algorithm 1 is very important because it provides a simple way to predict rhythms with higher accuracy. This algorithm can be adopted by any clinician/cardiologist and requires only the heartbeat type and the change of rhythm occurrence.

From the boxplot in Figure 3, it can be observed that RF algorithm does not provide surprisingly better results than a single DT model. But we can see that Specificity and Sensitivity measures have considerable variations, represented by the outliers. This highlights the importance of using RF, since it makes it possible to better explore the space of

models that can be learned from the data. Furthermore, there is a greater guarantee that the learned models will not be overfitted, due to the random bootstrap of the induction subsets.

The prediction errors (Table 9) show how the previous rhythm rAFIB implies rB or rT rhythms, but the models learned do not always capture this association. One possible explanation is that the sequence of rAFIB, rB or rT rhythms were not learned well, because there are many more rN rhythms preceded by rAFIB than rB or rT rhythms, according to rows 3, 4 and 5 on Table 1.

ROC curve (Figure 4) emphasizes the high performance of the proposed approach, since the mean AUC is around 0.99, and only for fold 5 it is 0.96.

Since the state-of-the-art methodologies cited in the Introduction section do not employ an approach similar to the one proposed here, comparisons are not feasible. It is noteworthy that they use directly the ECG signal or features generated from it. Besides that, none of these works propose the specific Bigeminy/Trigeminy classification.

Although the proposed approach has achieved higher performance, it has some limitations: (1) only the immediately posterior rhythm is predicted and (2) it is dependent on the mark of the rhythm change. Some of the works mentioned in the Introduction do not surpass these limitations. Lastly, in the next updates of the proposed approach (future works) these barriers must be overcome. In addition, it is also necessary to validate this approach in other databases.

DECLARATION OF COMPETING INTEREST

None declared.

FUNDING

This research did not receive any scholarship or funding from agencies that promote scientific research.

REFERENCES

- Ahn, M.S. (2013). Current concepts of premature ventricular contractions, *Journal of Lifestyle Medicine*, 3(1), 26–33.
- Amir, M., Mappangara, I., Kabo, P., Hasanuddin, Z., Setiadji, R. & Zam, S.M. (2020). Park Algorithm as Predictor of Premature Ventricular Contraction Origin in Three-Dimensional Mapping Electrophysiological Studies, *International Journal of General Medicine*, 13, 1083–1092. DOI: <https://doi.org/10.2147/IJGM.S275188>.
- Andersen, R.S., Peimankar, A. & Puthusserypady, S. (2019). A deep learning approach for real-time detection of atrial fibrillation. *Expert Systems with Applications*, 115, 465–473. DOI: <https://doi.org/10.1016/j.eswa.2018.08.011>.
- Awad, M. & Khanna, R. (2015). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, New York, Apress Open.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>.
- Bazi, Y., Hichri, H., Alajlan, N. & Ammour, N. (2013). Premature Ventricular Contraction Arrhythmia Detection and Classification with Gaussian Process and S Transform, in: *IEEE Fifth Int. Conference on Comput. Intell., Commun. Syst. and Networks*, 36–41. DOI: <https://doi.org/10.1109/CICSYN.2013.44>.
- Cappiello, G., Das, S., Mazomenos, E.B., Maharatna, K., Koulaouzidis, G., Morgan, J. & Puddu, P.E. (2015). A statistical index for early diagnosis of ventricular arrhythmia from the trend analysis of ECG phase-portraits, *Physiological Measurement*, 36(1), 107–31. DOI: <https://doi.org/10.1088/0967-3334/36/1/107>.
- Chugh, S.N. (2012). *Textbook of Clinical Electrocardiography for Postgraduates, Residents and Practicing Physicians*, 3 ed., London, Jaype.
- Fred, K. (2009). *ECG Interpretation: From Pathophysiology to Clinical Application*, Springer, New York.
- Garcia, T.B. & Miller, G.T. (2004). *Arrhythmia Recognition: The Art of Interpretation*, Jones and Bartlett Publishers, Burlington.
- Gerry, K. & Lemery, R. (2018). *Fast Facts: Cardiac Arrhythmias*, Health Press, Karger.
- Goodfellow, S.D., Goodwin, A., Greer, R., Laussen, P.C., Mazwi, M. & Eytan, D. (2018). Towards understanding ECG rhythm classification using convolutional neural networks and attention mappings, *Machine Learning for Healthcare*, 85, 83–101.
- Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.Ch., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K. & Stanley, H.E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23), e215–e220. DOI: <https://doi.org/10.1161/01.cir.101.23.e215>.
- Hajeb-Mohammadalipour, S., Ahmadi, M., Shahghadami, R. & Chon, K.H. (2018). Automated Method for Discrimination of Arrhythmias Using Time, Frequency, and Nonlinear Features of Electrocardiogram Signals, *Sensors* 18. DOI: <https://doi.org/10.3390/s18072090>.
- Hadia, R., Guldenring, D., Finlay, D.D., Kennedy, A., Janjua, G., Bond, R. & McLaughlin, J. (2017). Morphology-based detection of premature ventricular contractions, *Comput. Cardiol.* 44, 1–4. DOI: <https://doi.org/10.22489/CinC.2017.211-260>.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. DOI: <https://doi.org/10.1038/s41586-020-2649-2>.
- Hernández-Madrid, A., Lewalter, T., Proclermer, A., Pison, L. & Lip, G.Y.H. (2014). Remote monitoring of cardiac implantable electronic devices in Europe: results of the European Rhythm Association survey, *EP Europace*, 16(1), 129–132. DOI: <https://doi.org/10.1093/europace/eut414>.
- Ho, T.K. (1998). The random subspace method for constructing decision forests, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. DOI: <https://doi.org/10.1109/34.709601>.
- Latchamsetty, R. & Bogun, F. (2015). Premature ventricular complexes and premature ventricular complex induced cardiomyopathy, *Curr. Probl. Cardiol* 40, 379–422. DOI: <https://doi.org/10.1016/j.cpcardiol.2015.03.002>.
- Lee, H., Shin, S.Y., Seo, M., Nan, G.B. & Joo, S. (2016). Prediction of Ventricular Tachycardia One Hour before Occurrence Using Artificial Neural Networks. *Scientific Reports* 6. DOI: <https://doi.org/10.1038/srep32390>.
- Li, P., Liu, C., Wang, X., Zheng, D., Li, Y. & Liu, C. (2014). A low-complexity data-adaptive approach for premature ventricular contraction recognition, *Signal, Image Video Process.* 8, 111–120. DOI: <https://doi.org/10.1007/978-94-007-013-0478-6>.
- McKinney, W. (2010). Data structures for statistical computing in python, *Proceedings of the 9th Python in Science Conference*, 445, 56–61. DOI: <https://doi.org/10.25080/Majora-92bf1922-00a>.
- Moody, G.B. & Mark, R.G. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45–50. DOI:

- <https://doi.org/10.1109/51.932724>.
- Neubauer, S. (2007). The failing heart - an engine out of fuel, *The New England Journal of Medicine*, 356(11), 1140-1151. DOI: <https://doi.org/10.1056/nejmra063052>.
- Oliveira, B.R., Abreu, C.C.E. de, Duarte, M.A.Q. & Vieira Filho, J. (2019). Geometrical features for premature ventricular contraction recognition with analytic hierarchy process based machine learning algorithms selection, *Computer Methods and Programs in Biomedicine*, 169, 59-69. DOI: <https://doi.org/10.1016/j.cmpb.2018.12.028>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et. al. (2011). Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* 12, 2825-2830.
- Petruțiu, S., Sahakian, A.V. & Swiryn, S. (2007). Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans, *Europace* 9(7), 466-470. DOI: <https://doi.org/10.1093/europace/eum096>.
- Ricci, R.P., Morichelli, L., D'Onofrio, A., Calò, L., Vaccari, D., Zanolto, G., Curnis, A., Buja, G., Rovai, N. & Gargaro, A. (2013). Effectiveness of remote monitoring of CIEDs in detection and treatment of clinical and device-related cardiovascular events in daily practice: the HomeGuide Registry, *EP Europace*, 15, 970-977. DOI: <https://doi.org/10.1093/europace/eus440>.
- Taye, G.T., Shim, E.B., Hwang, H.J. & Lim, K.M. (2019). Machine Learning Approach to Predict Ventricular Fibrillation Based on QRS Complex Shape, *Frontiers in Physiology* 10. DOI: <https://doi.org/10.3389/fphys.2019.01193>.
- Teplitzky, B.A., McRoberts, M. & Ghanbari, H. (2020). Deep learning for comprehensive ECG annotation, *Rhythm*, 17(5), 881-888. DOI: <https://doi.org/10.1016/j.hrthm.2020.02.015>.
- Tom, F. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Tsai, C.H., Ma, H.P., Lin, Y.T., Hung, C.S., Huang, S.H., Chuang, B.L., Lin, C., Lo, M.T., Peng, C.K. & Lin, Y.H. (2020). Usefulness of rhythm complexity in heart failure detection and diagnosis. *Scientific Reports* 10. DOI: <https://doi.org/10.1038/s41598-020-71909-8>.
- Weaver, W.D., Cobb, L.A. & Hallstrom, A.P. (1982). Ambulatory Arrhythmias in Resuscitated Victims of Cardiac Arrest, *Circulation* 66(1), 212-218. DOI: <https://doi.org/10.1161/01.CIR.66.1.212>
- Wang, S., Cui, H., Ji, K., Zhu, C., Huang, X., Lai, Y. & Wang, S. (2021). Relationship Between Obstructive Sleep Apnea and Late Gadolinium Enhancement and Their Effect on Cardiac Arrhythmias in Patients with Hypertrophic Obstructive Cardiomyopathy. *Nature and Science of Sleep*, 13, 447-456. DOI: <https://doi.org/10.2147/NSS.S270684>.

To cite this paper:

Oliveira, B. R. de; Duarte, M. A. Q.; Vieira Filho, J. (2022). Early detection of Ventricular Bigeminy/Trigeminy rhythms. *Multi-Science Journal*, 5(1): 1-10. DOI: <https://doi.org/10.33837/msj.v5i1.1525>