33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Premature Ventricular Beat Classification Using a Dynamic Bayesian Network

Lorena S. C. de Oliveira, Rodrigo V. Andreão, Mario Sarcinelli-Filho

***Abstract*— This paper investigates the viability of using the dynamic Bayesian Network framework as a tool to classify heart beats in long term ECG records. A Decision Support System composed by two layers is considered. The first layer performs the segmentation of each heartbeat available in the ECG record, whereas the second layer classifies the heartbeat as Premature Ventricular Contraction (PVC) or Other. The use of both static and dynamic Bayesian Networks is evaluated through using the records available in the MIT-BIH database, and the results show that the Dynamic one performs better, obtaining 95% of sensitivity and 98% of positive predictivity, showing that to consider the temporal relation among events is a good strategy to increase the certainty about present events.**

I. Introduction

When solving certain problems, like in medical diagnosis, for instance, it is necessary to deal with the uncertainty inherent to the problem. Thinking on how to represent such uncertainty, the probability theory is a quite good tool, not only for allowing the use of random variables but also for its capability to express useful qualitative relations between beliefs and to process such relations to intuitively produce feasible conclusions.

This paper focuses on the study and development of the probabilistic methodology based on the Theorem of Bayes, the so called Bayesian Networks (BN). The case study to check the applicability of such technique is the detection of Premature Ventricular Contractions (PVC) in long term Electrocardiogram (ECG) records, obtained from the MIT-BIH Arrhythmia Database [1].

The use of probabilistic inference to develop intelligent systems has become a focus of interest to the academic community. In this work, such inference is accomplished using the BN, allowing getting conclusions as new information (new evidences) is analyzed. As an example, in medical diagnosis the physician can build a correct diagnostic based on the symptoms, which are his evidences. On the other hand, it is also possible to make inference to estimate non-observed variables.

Reference [2] evaluated the use of BN in the problem of heart beat classification, and the results obtained opened new perspectives on how to model uncertainty in such problem. Continuing such research, this work investigates the effect of taking into account the information of adjacent events, in a so called dynamic Bayesian network framework, making the probabilistic model closer to the way the physician analyses the ECG signal.

Manuscript received March 26, 2011. This work was partially supported by CAPES, a foundation of the Brazilian Ministry of Education and FAPEMIG, a foundation of the state of Minas Gerais, Brazil. L. S. C. Oliveira is with Science e Technology Institute (ICT), Federal University of Vales do Jequitinhonha e Mucuri, Teófilo Otoni, Minas Gerais, Brazil (phone: +55 33 3522-6037, lorenasco@yahoo.com.br). R. V. Andreão is with the Federal Institute of Education, Science and Technology of Espirito Santo, Vitoria, ES, 29040-780, Brazil (rodrigova@ifes.edu.br). M. Sarcinelli-Filho is with the Department of Electrical Engineering, Federal University of Espirito Santo, Vitoria, ES, 29075-910, Brazil (mario.sarcinelli@ele.ufes.br).

II. Premature Ventricular Contractions

The specific objective of this work is to check the feasibility of using BN to detect cardiac arrhythmias, taking the PVC arrhythmia as case study.

The term arrhythmia is used by physicians to mean any disturbance in the frequency, in the rhythm or in the propagation of the cardiac electrical impulses. The ECG offers a non-invasive and efficient way of diagnosing such different rhythmic disturbances of the heart, being, for such capability, the most common cardiac examination. Moreover, its long term version (a record containing thousands of heartbeats) has become usual.

Amongst the cardiac arrhythmias, PVC is the most common, especially when associated with acute myocardial infarction, although it can also occur associated with hypokalemia, digitalis intoxication and stress. Its greatest importance, however, is that it could be a precursor of more serious ventricular arrhythmias (ventricular extra systoles malignant), which give rise to tachycardia and ventricular fibrillation, the leading causes of cardiac arrest and death.

A heartbeat classified as PVC is characterized, in the ECG record, by a premature QRS complex followed, in most cases, by a compensatory pause. Another characteristic is the absence of the P wave before the premature QRS complex. Finally, a fourth characteristic of this arrhythmia is the deformation of the QRS wave, which becomes 'wider'. Such characteristics can be observed in the heartbeats of Figure 1, where the arrows shows PVC (the other are normal beats). Notice, specially, the deformed QRS complex and the absence of the P wave.

III. Dynamic Bayesian Networks

The BN is a method used to represent uncertain knowledge, once it allows arguing based on the theory of

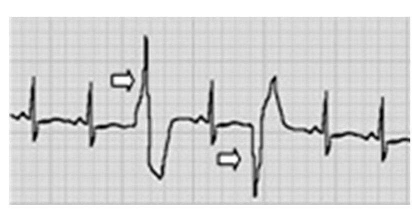

Fig 1: Polymorphic premature ventricular contractions.

probability [3]. In uncertain environments, however, it is necessary to continuously update its state to increase the certainty about it.

In general, the dependency relationships defined between the variables are static, although some applications demand a temporal relationship between the network nodes. Such temporal relationship is necessary when one wants to know the state of an environment that changes itself and receives new evidences along time.

Dynamic Bayesian Networks (DBN) have received increasing attention as a tool to model complex stochastic processes [4], because they manage to capture how the process evolves along time. A DBN is, essentially, a chain in which a classical BN is repeated as many times as necessary. Thus, this type of network models the probability distribution correspondent to a set of random variables $V_1,\cdots,V_N$, which can be split in observable and non-observable variables, generally referred to as E_t. Each variable in a DBN is associated to a time instant *t*, and is referred to as V_t^i, and the distribution of some variables V_t^i in the time instant *t* may depend on some variables in the time instant *t-1,* V_{t-1}^i, so that it is defined by a conditional distribution $P(V_t^i | V_{t-1}^i)$.

A DBN is a pair $D = (B_1, B_\rightarrow)$, where B_1 is the BN that defines an initial state (V_1) and $B_\rightarrow$ is a temporal BN representing two time instants (2TBN), defining $P(V_t^i | V_{t-1}^i)$ through a acyclic graphic representation characterized by [5]

$$P(V_t \mid V_{t-1}) = \prod_{i=1}^{N} P(V_t^i \mid pa(V_t^i)), \tag{1}$$

where V_t^i is the *i*-th node, in the time instant *t*, and $pa(V_t^i)$ are the parent nodes of V_t^i in the graph.

The variables can have different forms (like tables or Gaussian functions, for instance). By their turn, the parent nodes of a node V_t^i are the same, in the first time instant and in the subsequent ones as well. The arcs connecting nodes in different time instants go from the left to the right, thus representing the time going by. If there is an arc connecting a node V_{t-1}^i to a node V_t^i, such variable is called a persistent one. Consequently, one can say that when arcs connect two nodes in distinct time instants they model the persistence of a phenomenon along time. By their turn, arcs in the same time instant define an immediate causal relationship.

The semantic of the DBN can be defined extending the definition of 2TBN over an interval of T time instants. In such case, the joint probability distribution would be [5]

$$P(V_{i:T}) = \prod_{t=1}^{T} \prod_{i=1}^{N} P(V_t^i \mid pa(V_t^i)). \tag{2}$$

IV. METHODOLOGY

The database MIT-BIH *Arrhythmia Database* [1], used to test the system here designed to detect PVC, was split in two parts, one for training and the other for testing the network under analysis. The network training corresponds to the adjustment of its quantitative part, i. e., the adjustment of the probability tables. The test step, in turn, is used to check the credibility of the classification accomplished. For all tested cases, the network was implemented using the Bayesian Network Toolbox [6] of MATLAB®.

Figure 2 presents the architecture of the whole Decision Support System adopted to detect the PVC beats in long term ECG records. It is a two layers one: the first one, layer 0, corresponds to the segmentation of the beats, based on HMM, and is the same proposed in [7]. The second layer, layer 1, the focus of this work, uses a DBN as the classifier.

The topology of the DBN is shown in Figure 3. The arcs linking nodes in different time instants go from the left to right, and represent the time line. The choice of the number of time instants to be considered in connection with the DBN was based on tests performed with distinct network topologies, getting the conclusion that three time instants are enough for a good classification.

The random variables are described by the following nodes that constitute the DBN:

- '*PVC*', representing the possibility of occurrence for the arrhythmia being detected;
- '*Premature Beat*', indicating the probability of occurrence of a premature QRS complex;
- '*Ventricular Beat*', indicating the probability of occurrence of a wider QRS complex;
- *'LL'*, containing the numerical value of the likelihood associated to the QRS complex of the heartbeat being analyzed, used to determine the value of the node '*Ventricular Beat*'. Such node corresponds to one of the evidences delivered to the

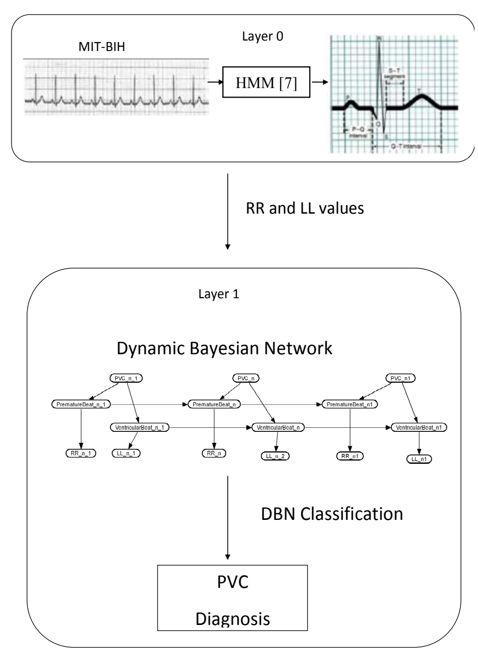

Fig 2: Architecture of the Decision Support System proposed to detect PVC beats, using the DBN.

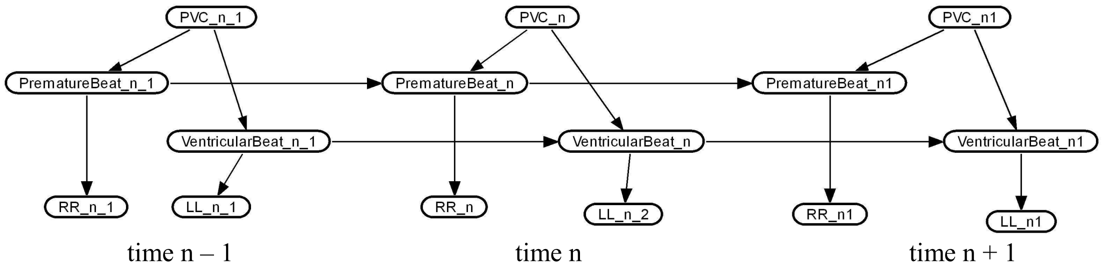

Fig 3: DBN considering three time instants.

classifier, and its value is calculated using the algorithm proposed in [7];

- *'RR'*, containing the numerical value of the time correspondent to the interval between two consecutive R peaks (the *RR* interval), used to determine the value of the node '*Premature Beat*'. Such node corresponds to the second evidence delivered to the classifier, and its value is also calculated using the algorithm in [7].

The quantitative representation associated to the nodes '*PVC*', '*Premature Beat*' and '*Ventricular Beat*' is described through a binary table, with the values True and False. The other nodes, '*LL*' and '*RR*', are quantitatively represented by Gaussian distributions, which are estimated through algorithms that generate histograms, using examples obtained from the training data set.

All the nodes are observable during the training step, so that the algorithm ML (Maximum Likelihood), available in the BNT toolbox [6], can be used.

Notice that the value of the node '*PVC_n_1*' (*n_1* means the previous time instant, or the previous beat), influences the node '*PVC_n*' (*n* means the current time instant or the current beat), which, by its turn, will influence the node *'PVC_n1*' (*n1* means the next time instant or the next beat), although not directly.

V. Validating the Results

The results here presented are evaluated using Contingence Tables (see Table I), because it allows a more general visualization of the errors and gives an effective measure of the quality of the classification model when showing the number of right classifications (true positive and true negative) and wrong classifications (false positive and false negative). Based on such table one can evaluate the system performance through the performance indexes

$$Sensitivity(Se) = \frac{a}{a+c}, \tag{3}$$

$$Specificity(Sp) = \frac{d}{b+d}, \tag{4}$$

$$PredictivePositiveValue(PPV) = \frac{a}{a+b}, \text{ and} \tag{5}$$

$$PredictiveNegativeValue(PNV) = \frac{d}{c+d}. \tag{6}$$

TABLE I
Representation of a Contingence Table for the classes 'PVC' or 'Other'

CONTINGENCE TABLE		
	LABEL IN THE DATABASE	
BN CLASSIFICATION	PVC	OTHER
PVC	*a* (TRUE POSITIVE)	*b* (FALSE POSITIVE)
OTHER	*c* (FALSE NEGATIVE)	*d* (TRUE NEGATIVE)

In this work, as the classification of the heartbeats is performed considering just two classes, the PVC beats and Other beats, and the class searched for is the PVC one, the precision of the classification will be evaluated by the indexes *Se* and *PPV*. In particular, the closer such indexes are to 100%, the better the performance of the classifier is. In such context, the confidence interval is also used, to define the interval containing such parameters [8]. For the *Se* index, the confidence interval is

$$\ell = \pm 1.96\sqrt{\frac{Se(100-Se)}{N}} \tag{7}$$

with a similar value defined for *PPV*.

VI. Results

Several tests were run, aiming at finding the ideal percentage of beats to compound the training data set, randomly selecting the percentage of beats for such data set, training the network and checking the results using the remaining beats as the test data set. The best result was obtained considering 80% of the available beats as the training set, and the remaining 20% as the test set (they are disjoint data sets).

The following step was to run the learning process for the BN. Then, as the final step, the classification capability of the system was checked using the test data set. For the sake of comparison, two results are presented: the first one is associated to the static BN, whereas the second one is associated to the DBN.

As mentioned before, a DBN, in the first time instant, is also a classical (static) BN. Thus, the topology of the static BN is the one presented in Figure 3, breaking the temporal relations, i. e., the topology of the static BN is the one corresponding to the time instant *n-1*, for instance. The

TABLE II
Contingence Table and results obtained with the static BN.

Contingence Table		
	Label 'PVC'	Label 'Other'
'PVC' Class	850	15
'Other' Class	97	13118
Sensitivity (Se)		89.76%
Specificity (Sp)		99.89%
Predictive Positive Value (PPV)		98.26%
Predictive Negative Value (PNV)		99.26%
Confidence Interval - Se (CI - Se)		± 1.93
Confidence Interval - PPV (CI - PPV)		± 0.8

results associated to such static BN are presented in Table II, while those associated to the DBN shown in Figure 3 are presented in Table III. In both cases the results correspond to the classification of PVC beats in the time instant *n*, which is the current one.

The result associated to the DBN corresponds to an improvement of about 6% in comparison with the results associated to the static BN, in terms of Sensitivity, meaning that the DBN classifies PVC beats better than the static BN. Notice that such improvement is meaningful, mainly when considering the narrow confidence intervals.

Another aspect in which the DBN differs from the static BN because the probability values associated to the DBN allow a better discrimination between classes. Table IV illustrates such difference: it shows 13 beats misclassified by both static and dynamic BN, during the test step. There the probabilities from which both networks misclassified the 13 beats as Other, against the labels of the specialists, which are all PVC, are shown.

From the table, one can see that the probabilities that the beats are Other, for the DBN, are much lower than those for the static BN, in general. This means that the DBN has more doubt about the classification, although deciding by the class Other.

VII. Conclusion

The objective of this paper, to investigate the applicability of the DBN in the detection of PVC heart beats recorded in long term ECG, was accomplished. The result is that BN are effective as a framework for such task. Actually, two possible BN instances can be adopted, namely the static and the dynamic BN. Considering long term ECG records got from a database, a comparison between the static BN and the DBN is performed, and the result is that the better results are associated to the DBN. Moreover, when both networks misclassify a heartbeat, the probability value associated to the DBN is, in general, lower than the one associated to the static BN.

Based on this last feature, our next step is to adopt minimum probability thresholds for the two classes considered (PVC and Other), such that if such thresholds are higher than the probability that the heartbeat being analyzed pertains to any of the possible classes the heartbeat is considered as a not classified one, and is put aside to be analyzed by a specialist. The idea, in this case, is to continue using the DBN, for the lower certainty it gives when misclassifying.

TABLE III
Contingence Table and results obtained with the DBN of Figure 3.

Contingence Table		
	Label 'PVC'	Label 'Other'
'PVC' Class	872	14
'Other' Class	45	13119
Sensitivity (Se)	95.09%	
Specificity (Sp)	99.86%	
Predictive Positive Value (PPV)	98.42%	
Predictive Negative Value (PNV)	99.66%	
Confidence Interval - Se (CI - Se)		± 1.4
Confidence Interval - PPV (CI - PPV)		± 0.8

TABLE IV
Probability values for 13 PVC beats misclassified as Other by the DBN and the static BN as well.

DBN (%)	Static BN (%)
75.38	95.45
83.40	96.89
91.88	96.26
81.41	76.55
74.18	95.10
64.58	93.43
88.78	89.39
90.89	96.15
73.70	94.89
83.57	95.30
91.29	96.28
88.01	89.86
71.00	94.45

References

[1] MIT-BIH. Arrhythmia DataBase Directory, viewed 14/03/2011, Available:http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm.

[2] L. S. de Oliveira, R. V. Andreão and M. Sarcinelli-Filho, Detection of Premature Ventricular Beats in ECG Records Using Bayesian Networks Involving the P-Wave and Fusion of Results, IEEE Eng Med Biol Soc. 2010;2010:1131-4.

[3] E. Castillo, J. M. Gutiérrez and A. S. Hadi, *Expert Systems and Probabilistic Network Model*s. Springer-Verlag, New York, 1997.

[4] A. Darwiche, "Constant-Space Reasoning in Dynamic Bayesian Networks," *International Journal of Approximate Reasoning*, vol. 26, no. 3, April 2001, pp 161-178.

[5] D. Belot, "Fusion de données avec des réseaux bayésiens pour la modélisation des systèms dynamiques et son application en telemedicine". Ph. D. dissertation, École doctorale IAE+M Loraine UFR STMIA, 2002.

[6] BNT, How to use the Bayes Net Toolbox (2007) viewed 14/03/2011. Available:http://www.cs.ubc.ca/~murphyk/Software/BNT/usage.html

[7] R. V. Andreão, B. Dorizzi and J. Boudy, "ECG Signal Analysis Through Hidden Markov Models," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, August 2006, pp. 1541-1549.

[8] R. Larson and B. Farber, *Elementary Statistics: Picturing the World*. Prentice-Hall, Englewood Cliffs, NJ, 2003.