

Kardio and Calicot: a comparison of two cardiac arrhythmia classifiers

E.Fromont, M.-O. Cordier and R.Quiniou

IRISA/INRIA, Campus de Beaulieu,
35042 Rennes Cedex FRANCE
{efromont,cordier,quiniou}@irisa.fr

A.I.Hernandez

LTSI, Campus de Beaulieu,
35042 Rennes Cedex FRANCE
Alfredo.Hernandez@univ-rennes1.fr

Abstract

This paper gives a comparison of two different systems that induce cardiac arrhythmia rules by symbolic learning: Kardio and Calicot. In particular, it proposes a detailed methodology to compare them and gives some results of this comparison.

Introduction

Coronary Care Units (CCU) were introduced in the 60's in order to monitor the vital functions of patients suffering a cardiac attack and, especially, to prevent, detect and control lethal arrhythmias by therapeutic actions. Cardiac arrhythmia detection and recognition have been studied in order to assist physicians and trigger alarms when necessary. In this article, we compare two systems that focus on this subject: Kardio [1] and Calicot [2]. Both systems can induce cardiac arrhythmia identification rules by symbolic learning. The aim of this paper is to give a methodology to compare these two different systems by a specific evaluation method that handle their differences while preserving a valid, quantitative comparison. The first part sketches the architectures and principles of the two systems. The second part presents the comparison methodology and the obtained results. The last part concludes on the positive features of each system and their possible future.

1 Compared architectures

1.1 Presentation

This section does not give a detailed description of each system architecture but points out their differences and similarities as shown in Figure 1. Further details can be found in [1] for Kardio and in [2] for Calicot.

The aim of Kardio is to diagnose cardiac arrhythmias from ECG descriptions. To do so, it looks for rules that describe all possible cardiac arrhythmias (single or multiple) corresponding to a given symbolic description of an ECG. Rule learning relies on a qualitative model of the heart that simulates the cardiac

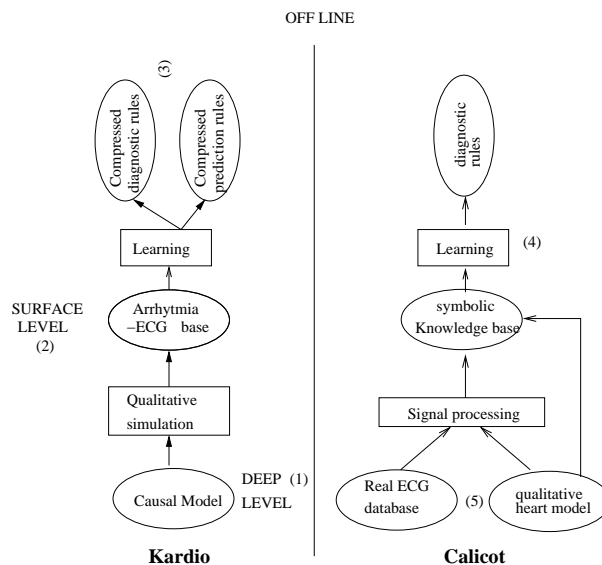


Figure 1: Architectures

electrical activity: over 2,400 heart disorders can be related to over 140,000 ECG descriptions (see number 1, Fig.1). Then, using deductive and inductive inference techniques, simulations produced by the qualitative model are automatically transformed into a set of compressed prediction and diagnostic rules (number 3, Fig.1). The diagnostic rules can answer the question “which heart disorders could be the source of a given ECG feature?”. They can be used for a precise diagnosis only using a diagnoser (ie an abductive method) and are difficult to compare with the expert system-like rules of Calicot. Thus, we have decided to focus on the prediction rules. They are causal rules of the form :

$P \Rightarrow (S_1 \vee S_2 \vee \dots \vee S_n)$ where P is an arrhythmia and S_1, \dots, S_n are selected ECG features of the form $S_i \equiv (S_{i_1} \wedge S_{i_2} \dots S_{i_n})$.

These rules are used to filter out non possible diagnosis: if a given ECG does not match the ECG description, the arrhythmia P is eliminated as a possible diagnoses. Kardio has been able to induce rules for 943 heart conduction defects from 5,240 ECG de-

scriptions.

The architecture of Calicot can be described in two steps : the first one, on which we focus, is done off-line (see Figure 1) and its aim is to build a set of high-level symbolic characterizations of cardiac arrhythmias, directly from real ECGs [2]. The learning step (number 4, Fig.1) relies on inductive logic programming (ILP) techniques. It makes use of learning examples (number 5, Fig.1) which are either real signals (like the labelled ECG signals from the MIT-BIH database [6]) or signals obtained by simulating arrhythmias on the Carmen cardiac model [5]. The second step is an on-line step which is in charge of analyzing the signal and identifying arrhythmias by matching the symbolic representation of the signal to prestored characterizations.

1.2 Analysis

In Calicot, the qualitative description of the ECG is computed from signal analysis methods. In contrast, the qualitative description of the signal in Kardio is directly given by the heart model (there is no signal processing step). In Calicot, the qualitative language is bounded by signal processing technologies because a very precise description of each heart wave is very difficult to obtain directly from a real signal. The language used in Kardio could be, thanks to the model, as rich and powerful as needed. For example, ambiguities remain in distinguishing between a Left Bundle Branch Block (LBBB) and a Right Bundle Branch Block (RBBB) on a real signal (both arrhythmias come from an intraventricular conduction disturbance but the first one comes from the left bundle branch and the second one from the right bundle branch). Kardio avoids this problem by using attribute values like *wide-lbbb* or *wide-rbbb* to describe the signal, even if this refinement level is difficult to reach by signal processing algorithms. We can then expect that the discrimination power of Kardio is better than the discrimination power of Calicot since Calicot cannot distinguish some arrhythmias.

Nevertheless, as explained in section 1.1, the final aim of Calicot is to analyze the signal on-line in order to identify arrhythmias by matching the symbolic representation of the signal to prestored patterns. Consequently, the qualitative language needs to fit what signal processing algorithms can currently achieve. Using Kardio rules for on-line analysis would mean to use signal processing algorithms able to produce on-line very detailed descriptions of the signal adapted to Kardio language. However, even if signal processing technologies are evolving very quickly, so precise descriptions are unconceivable in the next few years (especially under the noisy conditions which are often associated with CCU or ambulatory recording).

In contrast, for the same reasons, the validation of the Calicot rules can be done on real signals. Indeed, it is easy to translate the Prolog rules induced by the ILP module into chronicles [4] and, to compare the on-

line diagnosis results with labels provided by experts annotations. For Kardio, the only way to validate results is to ask an expert. We can then wonder if this is reliable. Indeed, a human expert judges the rules according to his own criteria and could not be able to evaluate information coming from an unknown sensor.

Finally, the semantics of the rules of the two systems is different. To give a precise diagnosis, it is not possible to use Kardio diagnostic rules without a diagnoser. However, the authors suggest to use their prediction rules by modus tollens, and then, to filter out the possible diagnoses by identifying that some symptoms are absent in the ECG description. On the contrary, Calicot induces associative rules which link symptom patterns to disorders and can be directly used for diagnosis by modus ponens.

2 Proposed comparison methodology

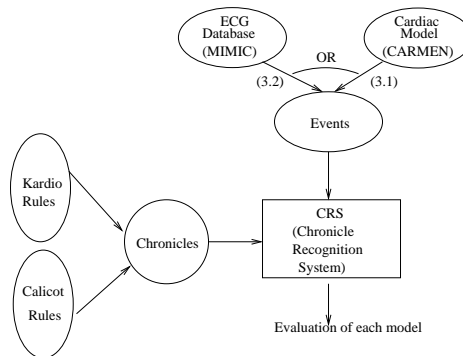


Figure 2: Comparison methodology

Figure 2 shows the principles of the methodology. To compare the two systems, we have selected a few rules for common arrhythmias from Kardio and Calicot : Sinus rhythm, LBBB and Mobitz rules. Each rule is then transformed into a CRS chronicle[4].

Chronicle recognition consists in skimming the flow of events coming from an observed process and detecting the specific events that belong to a chronicle. This process is similar to pattern-matching associated with temporal constraint satisfaction. In this article, chronicles are used to compare the detection capacity of each system from a set of events produced directly from real ECGs or from a cardiac model.

The next step is to translate Kardio prediction rules into CRS chronicles. Since we are dealing with a subset of Kardio knowledge, the closed world assumption is not valid and we cannot use the predicate completion [3] to transform the prediction rules into rules concluding positively on disorders. We have then decided to transform Kardio prediction rules by taking a weak version of the contrapositive. The resulting rule $(S_1 \vee S_2 \vee \dots \vee S_n) \Rightarrow \diamond P$ can be interpreted as follows: “if $(S_1 \vee S_2 \vee \dots \vee S_n)$ describes the ECG then P is a possible disorder”. This implies that if an arrhythmia is recognized, it does not mean that this arrhythmia

is the only possible diagnosis corresponding to a given ECG. We can just assume that a non-recognized arrhythmia is not a possible diagnosis.

The cardiac model used for the experiments is Carmen. Carmen is a macroscopic-level semi-quantitative cardiac model that is able to synthesize ECG signals and generate a physiological interpretation by means of ladder diagrams [5]. Different cardiac rhythm disorders can be simulated by manually defining an appropriate set of model parameters or by direct identification of the model parameters from real ECG signals. During a simulation, the model can also generate different symbolic representations of each synthesized ECG wave, describing its instant of occurrence, its morphology and its relation with the preceding wave(s). These symbolic representations have been constructed so as to be compatible with Kardio and Calicot description languages. Figure 3 shows a generated symbolic ECG in the Kardio and Calicot description language. These events are the input of CRS.

CALICOT	KARDIO
4377 qrs[abnormal]	4377 qrs[wide_LBBB]
5208 qrs[abnormal]	5208 qrs_ectopic[wide_LBBB]
5239 p_wave[normal]	5239 p_wave[normal]
6323 p_wave[abnormal]	6323 p_wave[abnormal]
6525 qrs[abnormal]	6525 qrs[wide_LBBB]
7408 p_wave[normal]	7408 p_wave[normal]
7618 qrs[abnormal]	7618 qrs[wide_LBBB]
8111 qrs[abnormal]	8111 qrs_ectopic[wide_LBBB]
8493 p_wave[abnormal]	8493 p_wave[abnormal]
.....

Figure 3: Example of an ECG description for Calicot and Kardio

The chronicle recognition results are then used to evaluate the recognition performance of each system. To cope with the differences between the semantics of the rules of the two systems, we have decided to count as a true positive recognition (TP) an arrhythmia which is in the set of possible diagnoses and should be recognized, as false negative (FN) an arrhythmia which is not in the set whereas it should have been, and as false positive (FP) an arrhythmia which is in the set whereas it should not be. In the FP case, if the arrhythmia which should be recognized is not one of the three studied arrhythmias, it is exceptionally count as a TP if the detected arrhythmia is the *sinus rhythm*. Indeed, in Kardio rules, the *sinus rhythm* can often be combined with other disorders and in this case, it is still possible that the right arrhythmia would have also been recognized. In practice, the true negative recognitions (TN) are computed as $TN = Tot - TP + FP + FN$ where Tot is the total amount of recognitions.

The criteria used for the comparison are the *sensitivity* which gives the probability of correct classification of a given observed rhythm, and the *specificity*,

that reflects the ability of the system to not propose a particular rhythm class if the observed rhythm does not belong to that class. They are respectively computed by :

$$SENS = \frac{TP}{TP+FN}, SPEC = \frac{TN}{FP+TN}.$$

3 Results

Two experimentations were achieved and the results are given in confusion matrices in Tables 1,2,3 and 4. The first experiment compares directly Kardio and Calicot and, the second one compares Calicot with a weakened Kardio. In each experiment, the matrix rows represent the detection and the columns represent the annotation given by Carmen in the first experiment and by the MIT database for the second one. The word UK is used for “Unknown” arrhythmias ie arrhythmias which are not considered in this paper (for example *rbbb* or *pvc*). The word NR is used for “non recognized” arrhythmias.

3.1 Kardio vs Calicot

In the first experiment, it was decided to leave the Kardio language unchanged and to compare it directly to Calicot. To do so, we need to use the same ECG but with two different symbolic representations to fit the two systems. Since we had no real ECG described in Kardio language, we used Carmen (see section 2) to produce different symbolic descriptions for the same ECG (see Fig. 3). The synthesis is performed in three steps:

Firstly, before starting the simulation, the cardiac model is initialized. A set of model parameters, which has been previously identified from real ECG signals and represents a given cardiac pathology (LBBB or a Mobitz rhythm), is loaded into Carmen.

Secondly, in order to generate different scenarii associated with the same cardiac disorder during the simulation, a model driver algorithm modifies randomly the following physiological model properties, every 4 seconds:

- Heart rate: from 40 to 190 beats per minute, using a uniformly distributed random variable.
- Atrio-ventricular conduction delay: a normally distributed random variable is used to define a conduction delay between 80 and 320 ms. This delay is distributed throughout the different atrio-ventricular structures of the model.
- Bundle branch conduction delay: the altered bundle branch (left or right) is chosen randomly and its conduction delay is defined with a normal distribution between 11 and 50 ms.
- Ectopic focus activation: an ectopic focus with a uniform random discharge period (defined between 1200 and 2100 ms) and a randomly chosen ventricular location is activated with probability 0.2.

Thirdly, at the end of the simulation, the internal symbolic representation of each wave generated by the model during the simulation, is translated into Calicot or Kardio language.

The rules learned by the two systems are then transformed into CRS chronicles. Examples of CRS chronicles for Kardio and Calicot are given in Figures 4 and 5. A comment is given after each event to give a brief description of the meaning of the predicate or, the Prolog rule from which the chronicle is generated. We can notice that only one chronicle is needed to recognize an LBBB with Calicot whereas the chronicle shown in Figure 5 is the first one of thirteen chronicles that describe the LBBB arrhythmia in Kardio. Indeed, Kardio language is a lot more precise than that of Calicot. For example, in Figure 5, we can see that the dominant QRS should be *wide_lbbb* and the ectopic QRS should be whether *wide_lbbb* or *wide_other* whereas in the Calicot rule shown in Figure 4, there is no difference between a dominant and an ectopic QRS, we only know that the shape of the QRS wave should be *abnormal*.

```

chronicle lbbb[]() {
occurs(0,0,p_wave[*] ,(start+1,R0-1))//no p_wave in [START,R0-1]
occurs(0,0,qrs[*] , (start+1,R0-1))//no qrs in [START, R0-1]
event(qrs[?w0], R0) //(qrs( R0 ,abnormal, _ ),
?w0 in {abnormal})

occurs(0,0,p_wave[*] ,(R0+1,P1-1))//no p_wave in [R0+1, P1-1]
occurs(0,0,qrs[*] , (R0+1,P1-1))//no qrs in [R0+1, P1-1]
event(p_wave[?w1], P1) //p_wav( P1 ,normal, R0 ),
?w1 in {normal}
R0 < P1

occurs(0,0,p_wave[*] ,(P1+1, R1-1))//no p_wave in [P1+1, R1-1]
occurs(0,0,qrs[*] , (P1+1, R1-1))//no qrs in [P1+1, R1-1]
event(qrs[?w2], R1) //qrs( R1,abnormal, P1),
?w2 in {abnormal}

P1 < R1
R1 - P1 in normalpr1 //pr1( P1 , R1 ,normal)

end - start in nb_cycles1}

```

Figure 4: A CRS chronicle for Calicot corresponding to the LBBB arrhythmia

The experiment results are given in Table 1 for Kardio and Table 2 for Calicot. First, we can notice that there are a lot of non recognized *mobitz* for both systems. This comes from the arrhythmia annotations provided by Carmen. Indeed, Carmen generates events randomly. It could generate some rare event patterns labeled as a *mobitz* (for example, four consecutive p-waves). However, since those patterns are not very common in medicine, the corresponding rules have not been induced by both systems and then, these patterns are not recognized. This brings a lot of false negative for the *mobitz* class. Moreover, Calicot rules for *mobitz* are more precise than Kardio. Indeed, it specifies that the p-wave occurring in a *mobitz* should be *normal* whereas Kardio does not specify anything on the shape of the p-wave so the latter has more recognitions for *mobitz* and his sensitivity

```

chronicle lbbb[]() {
occurs(0, 0,qrs[*] , (start+1, R0-1)) //no qrs in [START, R0-1]
event(qrs[?w0], R0) //qrs(R0,_,wide_LBBB),
?w0 in {wide_LBBB}

occurs(0, 0,qrs[*] , (R0+1, R01-1)) //no qrs in [R0+1, R01-1]
event(qrs_ectopic[X], R01) //qrs_ectopic(R01,R0,X),
R0 < R01

X in {wide_LBBB,wide_other}
occurs(0, 0,qrs[*] , (R01+1, R1-1)) //no qrs in [R01+1, R1-1]
event(qrs[?w3], R1) //qrs(R1,_,wide_LBBB),
?w3 in {wide_LBBB}

R0 < R1
R1 - R0 in shorttrr1 //rr1(R0, R1, short)

end - start in nb_cycles1}

```

Figure 5: A CRS chronicle for one of the Kardio rule for LBBB

	mobitz	lbbb	normal	UK	Total
mobitz	114	0	0	32	146
lbbb	0	20	0	0	20
normal	14	6	2765	0	2785
NR	909	3	0	0	912
Total	1037	29	2765	32	3863
Sensit	0.11	0.69	1	0	
Specif	0.99	1	0.98	1	

Table 1: The confusion matrix for Kardio rules with Carmen signal

	mobitz	lbbb	normal	UK	Total
mobitz	30	0	0	20	50
lbbb	0	22	0	705	727
normal	1	0	1816	0	1817
NR	1328	7	0	0	1335
Total	1358	29	1816	725	3928
Sensit	0.02	0.76	1	0	
Specif	0.99	0.66	1	1	

Table 2: The confusion matrix for Calicot rules with Carmen signal

is better. Besides, we can notice that the results for *lbbb* and more particularly the number of false positive is a lot better for Kardio (0) than for Calicot (705). This comes from the fact that Calicot never makes the difference between an *rbbb* and an *lbbb* as explain in section 1.2. Finally, there are a lot more TP in the *normal* class for Kardio (2785) than for Calicot (1816). This comes from the choice we made about the detection of unknown arrhythmias as explained in section 2. Indeed, when Kardio detects a normal rhythm instead of an unknown one (for example *rbbb*) we have considered that it was a correct detection for the *normal* class because Kardio has eliminated the *mobitz* and the *lbbb*.

	mobitz	lbbb	normal	UK	Total
mobitz	427	0	0	0	427
lbbb	0	2006	8	1106	3120
normal	0	0	2292	0	2292
NR	0	0	291	0	291
Total	427	2006	2591	1106	6130
Sensit	1	1	0.88	0	
Specif	1	0.73	1	1	

Table 3: The confusion matrix for Calicot rules

	mobitz	lbbb	normal	UK	Total
mobitz	427	0	0	0	427
lbbb	0	2006	0	1432	3438
normal	0	0	7406	0	7406
NR	0	0	0	0	0
Total	427	2006	7406	1432	11271
Sensit	1	1	1	0	
Specif	1	0.85	1	1	

Table 4: The confusion matrix for weakened Kardio rules

3.2 Weakened Kardio vs Calicot

In a second step, we have weakened Kardio language to fit current signal processing algorithm possibilities. Every shape that was not described as *normal* was assumed *abnormal* and every ectopic QRS was considered as a dominant QRS. Indeed, nowadays, it is still difficult to differentiate an ectopic QRS from the dominant one or to feature precisely a wave shape just by analyzing the signal.

In this experiment, the signal comes from a real ECG and the symbolic description is the same for the two systems. An example of a CRS chronicle for the weakened Kardio is given in Figure 6. This chronicle corresponds to the same rule that was used to create the chronicle of Figure 5.

```
chronicle lbbb[]() {
occurs(0, 0,qrs[?],(start+1, R0-1)) //no qrs in [START, R0-1]
event(qrs[?w0], R0) //qrs(R0,_,abnormal),
?w0 in {abnormal}

occurs(0, 0,qrs[?],(R0+1, R1-1)) //no qrs in [R0+1, R1-1]
event(qrs[?w1], R1) //qrs(R1,_,abnormal),
?w1 in {abnormal}

occurs(0, 0,qrs[?],(R1+1, R2-1)) //no qrs in [R1+1, R2-1]
event(qrs[?w2], R2) //qrs(R2,_,abnormal),
?w2 in {abnormal}

R0 < R2
R2 - R0 in shortrr1 //rr1(R0,R2,short)

end - start in nb_cycles2 }
```

Figure 6: A CRS chronicle for one of the weakened Kardio rule for LBBB

Results are given in Tables 3 and 4. We can notice that the results are good and quite the same for Kardio and Calicot except that Kardio has twice as much

detections than Calicot. Indeed, the choices made to classify Kardio detections are all in Kardio advantage because if there is a multiple detection in Kardio, it is counted as a true positive for *normal* if the right solution is in the set of detected arrhythmias. In particular, for each *normal* rhythm, the weakened Kardio has detected the *normal* rhythm and a lot of other arrhythmias (*lbbb* or *mobitz*). We can also notice that there is a lot of FP for *lbbb* for both systems. This comes from the fact that Kardio can not distinguish between a *rbbb* and a *lbbb* with its new weakened language. These results are similar to those of Calicot presented in Table 2 for *lbbb*.

Conclusion

This paper has given a comparison of two cardiac arrhythmia classifiers with very different nature. It has proven to be very difficult due to the different semantics attributed to the rules of the two systems. We also have presented a qualitative evaluation methodology of Kardio which has never been done before. The comparison has shown that Kardio with its powerful language is more precise than Calicot whereas the latter is more adapted to current signal processing technologies and so, to an on-line application. As the signal processing techniques evolve, we plan to improve the knowledge base for the ILP module of Calicot to make Calicot language more powerful. An another interesting experiment will then be compare the new rules induced by Calicot with the rules of Kardio.

References

- [1] I. Bratko, I. Mozetič, and N. Lavrač. *Kardio: A Study in Deep and Qualitative Knowledge for Expert Systems*. MIT Press, Cambridge, MA, 1989.
- [2] G. Carrault, M-O. Cordier, R. Quiniou, and F. Wang. Temporal abstraction and inductive logic programming for arrhythmia recognition from ECG. *Artificial Intelligence in Medicine*, 28:231–263, 2003.
- [3] K. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*. Plenum Press, New York, NY, 1978.
- [4] C. Dousson, P. Gaborit, and M. Ghallab. Situation recognition: representation and algorithms. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 166–172, Chambry, France, 1993. Morgan Kaufman.
- [5] A. I. Hernández, G. Carrault, F. Mora, and A. Bardou. Model-based interpretation of cardiac beats by evolutionary algorithms: signal and model interaction. *Artificial Intelligence in Medicine*, 23(211-235), 2002.
- [6] G. B. Moody. Ecg database programmer’s guide. Harvard-MIT Division of

Health Sciences and Technology Biomedical Engineering Center, Ninth Edition, 1997.
<http://www.physionet.org/physiobank/database/mitdb/>.