Using fuzzy automata to diagnose and predict heart problems

Azahara Camacho, Mercedes G. Merayo, Manuel Núñez
Departamento Sistemas Informáticos y Computación
Universidad Complutense de Madrid, Spain
email: {mariaazc,mlmgarci,manuelnu}@ucm.es

***Abstract*—In this paper we introduce a formalism to specify the behavior of biological systems. Our formalism copes with uncertainty, via fuzzy logic constraints, an important characteristic of these systems. We present the formal syntax and semantics of our variant of fuzzy automata. The bulk of the paper is devoted to present an application of our formalism: a formal specification of the heart that can help to detect abnormal patterns of behavior. Specifically, our model analyzes the heartbeats per minute and the longitude of the RR waves of a patient. The model takes into account the age and gender of the patient, where age is considered to be a *fuzzy* parameter. Finally, we use real data to analyze the reliability of the model concerning the diagnosis and prediction of potential illnesses.**

***Index Terms*—modelling and simulation of biological systems; formal methods; fuzzy systems.**

I. Introduction

The use of formal approaches during the development of computer systems improves their reliability. Unfortunately, formal validation and verification activities require a deep knowledge of non-trivial algorithms and notations. Therefore, it will be difficult to obtain a wide use of model-based testing [1], [2] and model checking [3] in fields unrelated to formal methods. However, formal methods should guide the development of systems in the same way as blueprints guide the construction of buildings [4].

Although the use of *nature-inspired* methods in healthcare and medicine seems natural [5], [6], the few applications of formal methods to the field are not completely satisfying. The main problem is that existing formalisms did not adapt well. Consider the existing models of the heart, the main topic of this paper. Despite the use of formal approaches [7], [8], [9], [10], these formalisms do not take into account uncertainty and imprecision, something vital when dealing with biological systems. Therefore, we have inaccurate models and it is very difficult to obtain a good diagnosis using a bad model. For example, a model indicates the beginning of an arrhythmia when the duration of the RR waves is less than 934 milliseconds for patients between 30 and 39 years because a poor progression in this wave tends to be attributed to anterior myocardial infarction. If a patient is 35 years old and has a value equal to 934.05 milliseconds, then it will not be identified as a potential problem. This minimal distance to the limit should not be an impediment to analyze the evolution of the heart of this patient because, as in this case, a lack of treatment could have fatal consequences. Therefore, there is a need to develop new formal methods that can produce more appropriate *blueprints*. In this line, *fuzzy logic* is the perfect complement of classical formal methods. In fact, there are many proposals to define fuzzy automata [11], [12], [13], [14]. We introduce a simplified language because we do not need the whole expressive power of these formalisms and additional features would obscure the presentation of our models.

In this paper we are interested in modeling the behavior of the heart concerning the prediction of arrhythmias from the data derived from electrocardiograms: heartbeats and RR wave durations. Having a formal model of a system, even a biological system, allows us to automatically analyze data in order to detect possible failures. In particular, this model can be used to improve the accuracy of prediction and diagnosis processes. We introduce a formal model of the heart that can be used to detect and predict *erroneous* behaviors, giving to the healthcare professional additional information about possible premature patterns of suspicious heart routines. In order to assess the usefulness of the model, we will analyze real patients data to check whether our model detects existing illnesses.

The heart is the *engine* of the body with a pump-blood rate of approximately 5 liters per minute. Its physiognomy is characterized by four chambers: right/left atria and ventricles. The main functionality of these elements is to collect the deoxygenated blood from the body in the right atrium and then, pump it into the lungs via the right ventricle. Thanks to the lungs, blood is purified with oxygen, it passes through the left atrium and finally, it goes to the left ventricle where it is pumped to the rest of the body.

The heart generates electric stimulus about 60-100 heartbeats per minute by the sinus node. This is considered the correct level of stimulus for a healthy person. A rate out of this range can be produced by some cardiology pathology. The consequences of this situation are irregular levels of blood pressure, weakness and fatigue. Depending on the number of irregularities that are detected in a patient, they can be the preamble of one of the following arrhythmias: bradycardia or tachycardia. In the first case, the patient presents a slow heart rate followed by an insufficient blood supply. The latter one is the responsible of the majority of fatal arrhythmias and it is detected because of a fast heart rate and impair hemodynamics [15].

Due to the complexity of this tissue, there are numerous factors that have to be taken into account in order to analyze the level of gravity of a cardiac disease [16]. The main tool used for these diagnostics is the ECG (electrocardiogram): this is a process consisting in registering the electrical signs of the heart over a period of time using a set of electrodes, the majority of them being placed near the chest of the patient. This is the most popular noninvasive cardiology test and gives a general vision of the current structure and functions of the heart. The conventional ECG is the 12-lead ECG which is composed by 10 leads, three bipolar and six unipolar: I Lateral, II Inferior, III Inferior, aVR, aVL Lateral, aVF Inferior, V1 Septal, V2 Septal, V3 Anterior, V4 Anterior, V5 Lateral and V6 Lateral [17], [18]. In this work, we focused on the information provided by lead II, V1, V2, V4 and V5.

The interpretation of the signals of the ECGs is vital for understanding the behavior of the heart and the possible patterns of heart disease that they reflect. Similarly to other medical tests, the normal levels of ECGs are derived from the study of numerous patients (both healthy and with some suspicious pathology) [19], [20]. However, even with the availability of these big banks of data, there are patients that have not received a correct diagnosis of their heart disease. As a consequence, they developed a more serious illness or even their death [21], [22]. In order to avoid these cases, the research community have tried to improve the diagnosis process with the application of novel techniques. Some evidences of these attempts are the use of neural networks for prediction [23], data mining [24] and the comparison of different prediction techniques with the aim of detecting the best option for healthcare professionals [25]. Fuzzy logic has also played an important role in the effort to enhance the diagnosis of heart diseases [26], [27]. However, and this is an important drawback that we fix with our work, previous approaches do not combine different patient data with the aim of obtaining a more accurate diagnosis. In fact, there are different healthcare studies with evidences of successful results derived from data combination [28], [29]. This advantage can be obtained by combining data in two different ways: merging initial data with previous references of other patients and merging initial data with more information from the patient. According to this evidence, we decided to put in practice the same idea for predicting heart problems. We apply fuzzy automata to predict the unexpected behavior of the heart in some special classes of patients, including patients diagnosed with premature aging of heart tissues, sport people and patients with chronic illnesses. We combine data of the patient in order to find previously unknown patterns associated with possible undetected heart illnesses.

The rest of the paper is structured as follows. In Section II we introduce some basic concepts associated with the definition of systems using fuzzy logic. In Section III we present the syntax and semantics of our formalism, a variant of fuzzy automata. In Section IV, the bulk of the paper, we present our model of the heart and evaluate its usefulness with real data. Finally, in Section V we present our conclusions and some lines for future work.

II. Fuzzy logic: basic concepts

In this section we present some basic concepts of fuzzy logic that will be used in the next sections of the paper. First, let us note that the fuzzy framework does not strictly categorize statements to be true or false. On the contrary, in order to evaluate the truth or falsehood of an assessment, we consider a range of values in the interval $[0,1]$: the larger the value, the more confidence we have on the truthfulness of the statement. In this paper we only consider positive values. Thus, the relations that we consider will be defined over the set of non-negative real numbers, denoted by $\mathbb{R}_+$. Therefore, a fuzzy relation is a mapping from the Cartesian product $\mathbb{R}_+^n$ into the interval $[0,1]$.

Definition 1 *A* fuzzy relation $\overline{A}$ *is a function* $\overline{A} : \mathbb{R}_+^n \longmapsto [0,1]$. *Let* $x \in \mathbb{R}_+^n$. *We say that* $\overline{x}$ *is* not included *in* $\overline{A}$ *if* $\overline{A}(\overline{x}) = 0$. *We say that* $\overline{x}$ *is* included *in* $\overline{A}$ *if* $\overline{A}(\overline{x}) > 0$. *We say that* $\overline{x}$ *is* fully included *in* $\overline{A}$ *if* $\overline{A}(\overline{x}) = 1$. *The* kernel *of* $\overline{A}$ *is the set of elements that are fully included in* $\overline{A}$.

In order to provide a credibility threshold, it is necessary to define the concept of α-cut: in a fuzzy relation $\overline{A}$, we accept a relation for $\overline{x}$ if $\overline{A}(\overline{x})$ is above the threshold.

Definition 2 *Let* $\overline{A} : \mathbb{R}_+^n \longmapsto [0,1]$ *be a fuzzy relation and* $\alpha \in [0,1]$. *We define the* α*-cut of* $\overline{A}$, *written* $\mathtt{cut}_\alpha(\overline{A})$, *as* $\mathtt{cut}_\alpha(\overline{A}) = \{\overline{x} \in \mathbb{R}_+^n \mid \overline{A}(\overline{x}) \geq \alpha\}$.

Next we introduce some fuzzy relations that we use in this work and that are variations of the usual order relations between real numbers. For example, the next relation $\overline{x = y}^\delta$ holds, for a given δ, if the distance between x and y is *small*, formally, if the difference between them is smaller than δ. However, we quantify how similar these values are by not simply providing a binary true/false answer but a quantification of the distance.

Definition 3 *Let* $\delta \in \mathbb{R}_+$. *We define the three relations* $\overline{\cdot = \cdot}^\delta, \overline{\cdot \leq \cdot}^\delta, \overline{\cdot \geq \cdot}^\delta : \mathbb{R}_+^2 \longmapsto [0,1]$ *as follows:*

$$\overline{x = y}^\delta = \begin{cases} 0 & \text{if} \quad x \leq y - \delta \\ 0 & \text{if} \quad x > y + \delta \\ 1 + \dfrac{x-y}{\delta} & \text{if} \quad y - \delta < x \leq y \\ 1 - \dfrac{x-y}{\delta} & \text{if} \quad y < x \leq y + \delta \end{cases}$$

$$\overline{x \leq y}^\delta = \begin{cases} 1 & \text{if} \quad x < y \\ 0 & \text{if} \quad x > y + \delta \\ 1 - \dfrac{x-y}{\delta} & \text{if} \quad y \leq x \leq y + \delta \end{cases}$$

$$\overline{x \geq y}^\delta = \begin{cases} 1 & \text{if} \quad x > y \\ 0 & \text{if} \quad x \leq y - \delta \\ 1 + \dfrac{x-y}{\delta} & \text{if} \quad y - \delta < x \leq y \end{cases}$$

In addition to the previous three standard fuzzy relations, we introduce a ternary fuzzy relation to estimate whether a *given value fuzzily states between the other two. Formally, we define the relation* $\overline{\cdot \leq \cdot \leq \cdot}^{\delta} : \mathbb{R}_{+}^{2} \longmapsto [0,1]$ *as follows:*

$$\overline{z \leq x \leq y}^{\delta} = \begin{cases} 1 & \textit{if} \quad z < x < y \\ 0 & \textit{if} \quad x \leq z - \delta \vee x > y + \delta \\ 0 & \textit{if} \quad y < z \\ 1 + \dfrac{x - z}{\delta} & \textit{if} \quad z - \delta < x \leq z \\ 1 - \dfrac{x - y}{\delta} & \textit{if} \quad y \leq x \leq y + \delta \end{cases}$$

In the last fuzzy relation, let us note that if $z \leq y$ does not hold then $\overline{z \leq x \leq y}^{\delta}$ is equal to zero regardless the values of x and δ.

After these definitions, we need a method to check properties and value their degree of similarity. This can be fulfilled with the concept of the binary operation *triangular norm* (usually called t-norm).

Definition 4 *A* t-norm *is a function* $T : [0,1] \times [0,1] \longmapsto [0,1]$ *satisfying the following properties:*

- *Commutativity:* $T(\delta_1, \delta_2) = T(\delta_2, \delta_1)$.
- *Monotonicity: if* $\delta_1 \leq \delta_3$ *and* $\delta_2 \leq \delta_4$ *then* $T(\delta_1, \delta_2) \leq T(\delta_3, \delta_4)$.
- *Associativity:* $T(\delta_1, T(\delta_2, \delta_3)) = T(T(\delta_1, \delta_2), \delta_3)$.
- 1 *is the identity element:* $T(\delta, 1) = \delta$.
- 0 *is nilpotent:* $T(\delta, 0) = 0$.

Since t*-norms are commutative and associative, we will sometimes use* $T\{\delta_1, \delta_2, \delta_3, \ldots, \delta_r\}$ *as a shorthand of* $T(\delta_1, T(\delta_2, T(\delta_3, \ldots, \delta_r) \ldots))$.

We denote by $\triangle$ a generic t-norm. Abusing the notation, we will also use $\triangle$ to denote the function associated to $\triangle$. There are many t-norms in the literature. Among the most used we can mention the following:

1) Łukasiewicz t-norm: $T(\delta_1, \delta_2) = max(0, \delta_1 + \delta_2 - 1)$. We represent this t-norm with the symbol $\curlywedge$.
2) Gödel t-norm: $T(\delta_1, \delta_2) = min(\delta_1, \delta_2)$. We represent this t-norm with the symbol $\barwedge$.
3) Product t-norm: $T(\delta_1, \delta_2) = \delta_1 \cdot \delta_2$ (real number multiplication). We represent this t-norm with the symbol $\bigstar$.
4) Hamacher product t-norm: $T(\delta_1, \delta_2) = \dfrac{\delta_1 \cdot \delta_2}{\delta_1 + \delta_2 - \delta_1 \cdot \delta_2}$. We represent this t-norm with the symbol $\divideontimes$.

III. Formalization of fuzzy automata

As we have previously commented, there are several approaches in the literature to combine fuzzy logic and automata. In this paper we consider a variant, actually a simplification, of our previous work [14], [30] because it is closer to the widely used timed automata [31] theory and we can implement our approach in existing frameworks such as UPPAAL [32]. Next, we introduce some concepts that will be used to define our automata.

Definition 5 *We assume that we have a fixed finite set of actions, denoted by* Acts*. We will distinguish between* inputs*, preceded by* $?$*, and* outputs*, preceded by* $!$*. We will assume that we have a finite set of fields, denoted by* Fields*. Since this set is finite, we will assume that* $|\text{Fields}| = n$*, for a certain* $n \in \boldsymbol{N}$*, and that there is an implicit sorting of the fields so that* $\text{Fields} = \{f_1, f_2, \ldots, f_n\}$.

Actions will be used to indicate reactions of a system. We will use inputs when the automaton is receiving information from the *environment*. The environment usually is the system providing information about the patient. In our case study, inputs will be used, for example, to receive information about the gender of the patient or about the current BPM (heartbeats per minute). We will use outputs when we want to sent a message to the environment. For example, we can issue an alarm indicating that a potential problem has been found or we can tell the environment that it should start sending information about RR waves. Besides, fields will be used to store information, either received from or sent to, of the environment.

Our automata will have, as usual, states and transitions (identified with the edges of the underlying graph). In order to trigger a transition we need to perform the associated action, similar to classical automata, and check that a certain constraint, on the values of some of the fields, holds. In our setting, we consider a fuzzy interpretation of *holding*. Thus, we need to introduce the concept of *fuzzy constraint* and the evaluation of such a constraint.

Definition 6 *A* fuzzy constraint *is a formula built from the following BNF.*

$$C ::= \texttt{True} \mid C_1 \triangle C_2 \mid \overline{f \circledast r_1}^{\delta} \mid \overline{r_1 \leq f \leq r_2}^{\delta}$$

where $\triangle$ *is a* t*-norm,* $\circledast \in \{\leq, =, \geq\}$*,* $f \in \text{Fields}$*, and* $\delta, r_1, r_2 \in \mathbb{R}_+$*. We denote the set of fuzzy constraints by* $\mathcal{FC}$.

Let C *be a fuzzy constraint and* $\overline{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$ *be such that for all* $1 \leq i \leq n$ *we have that* x_r *is the current value of* f_r*. We inductively define the* satisfaction degree or grade of confidence (GoC) of C for $\overline{x}$*, written* $\mu_C(\overline{x})$*, as:*

$$\mu_C(\overline{x}) = \begin{cases} 1 & \textit{if} \quad C = True \\ \overline{x_r \circledast r_1}^{\delta} & \textit{if} \quad C = \overline{f_r \circledast r_1}^{\delta} \\ \overline{r_1 \leq x_r \leq r_2}^{\delta} & \textit{if} \quad C = \overline{r_1 \leq f_r \leq r_2}^{\delta} \\ \triangle(\mu_{C_1}(\overline{x}), \mu_{C_2}(\overline{x})) & \textit{if} \quad C = C_1 \triangle C_2 \end{cases}$$

Let us note that if $\delta = 0$ then fuzzy constraints become *usual* constraints. In this case, we will omit the bar and the 0 over the constraint. For example, we will write $f \leq r$ instead of $\overline{f \leq r}^{0}$. In the previous definition, let us remark that $\mu_C(\overline{x}) \in [0,1]$. We consider a reduced set of fuzzy constraints because these are the ones that we use in this paper. It is trivial to enrich the previous set by adding cases to the BNF introduced in the previous definition. Next we give an example where we show the usual kind of constraints that we will use in the automata appearing in this paper.

Age	Fuzzy Constraint	Grade of Confidence
16	$\overline{age \leq 19}^4$	1
22	$\overline{age \leq 19}^4$	0.25
85	$\overline{age \geq 80}^{16}$	1
77	$\overline{age \geq 80}^{16}$	0.8125
77	$\overline{60 \leq age \leq 69}^{13}$	0.3846
65	$\overline{60 \leq age \leq 69}^{13}$	1
55	$\overline{60 \leq age \leq 69}^{13}$	0.6154

Table I
EXAMPLES OF FUZZY CONSTRAINTS WITH AGES

Example 1 *One of the fields that we use in Section IV is the* age of the patient*. We will categorize a patient as a child, young, adult or elder. Consider a* 55 *years old patient whose heart has got prematurely elder due to his arduous work experiences. A naïve, deterministic approach would probably classify him as an adult because of his age, but according to the consequences of his past events, he also needs to be classified as an elder patient. In this kind of scenarios, it is necessary the application of fuzzy constraints that consider these situations.*

Applying this idea and according to the data set that we use in our work [19], we can classify patients in 8 different groups depending on his age: $16-19$, $20-29$, $30-39$, $40-49$, $50-59$, $60-69$, $70-79$ *and* $80-89$. *However, with the aim of not excluding any patient in our automata, we changed the limits in the first and last group of patients. These changes are applied with the fuzzy constraints* $\overline{age \leq 19}^4$ *and* $\overline{age \geq 80}^{16}$. *We decided to manage children and teenagers using the first fuzzy constraint. We set* $\delta = 4$ *so that we consider patients 4 years over that limit with a GoC under* 100%*: this grade is indirectly proportional to the number of years over 19. The second fuzzy constraint reflects the same idea, but with elderly patients over 80 years.*

Another fuzzy constraint that we use in our automata is $\overline{60 \leq age \leq 69}^{13}$. *In principle, the age of the patient has to be between* 60 *and* 69. *Since we have* $\delta = 13$, *we would also include in this group patients between* 47 *and* 72 *years. The confidence that we will have in the obtained results will decrease as the age approaches these extended limits, being strictly below* 1 *as soon as the age does not belong to the original interval. In order to show how these grades of confidence vary, in Table 1 we give some ages of patients and how they are classified according to these three fuzzy constraints, including the case of the patient previously mentioned. We do not show confidences equal to zero.*

These fuzzy constraints are taken from our automaton used to analyze real patients data. The extended bounds provided by δ *are set according to the pathologies that patients can develop depending on their age. In Section IV, we explain how the limits and the corresponding* δ *were established for the rest of fields.*

Now we are able to define the components that define a fuzzy automaton.

Definition 7 *A* fuzzy automaton *is a tuple* (S, s_0, T) *where:*

- S *is a finite set of locations.*
- s_0 *is the initial location.*
- $T \subseteq S \times \text{Acts} \times \mathcal{FC} \times S$ *is the set of edges. We write* $s \xrightarrow{a,C} s'$ *whenever* $(l, a, C, l') \in T$.

As we already said, fuzzy automata are directed graphs where transitions have an associated condition. Given a transition $(s, a, C, s') \in T$, if the automaton is in state s and receives/sends from/to the environment a pair $(a, \overline{x})$, where a is an action and $\overline{x}$ is a tuple including a value for each field, then the previous transition can be triggered if $\mu_C(\overline{x}) > 0$ and the automaton will move to state s'. Let us remark that if we receive from the environment, then a is an input and it is preceded by ?; otherwise, a is an output and it is preceded by !. Inputs have priority over outputs: the system should prioritize receiving additional data over sending a message. We will formalize these ideas in the following definition.

Next, we are going to define the operational behavior of fuzzy automata in order to obtain their (fuzzy) traces. Therefore, we start in the initial state of the automaton, produce pairs (action, values) and trigger a transition labelled by the action if the values, associated with the fields, are included in the fuzzy relation induced by the constraint. We decorate transitions with a real number $\epsilon \in [0, 1]$ indicating its certainty. First, we define a single transition and then we concatenate transitions to conform traces.

Definition 8 *Let* $A = (S, s_0, T)$ *be a fuzzy automaton and* $\triangle$ *be a* t*-norm. Given states* $s_1, s_2 \in S$, *we have a* transition *from* s_1 *to* s_2, *after performing the action* $a \in \mathcal{A}ct$ *for the values* $\overline{x} \in \mathbb{R}_+^n$ *with confidence* ϵ, *denoted by* $s_1 \xrightarrow{(a,\overline{x})}_\epsilon s_2$, *if the following conditions hold:*

- *There exists* $C \in \mathcal{FC}$ *such that* $(s_1, a, C, s_2) \in T$.
- $\mu_C(\overline{x}) = \epsilon$ *and* $\epsilon > 0$.
- *If* a *is an output, then there do not exist an input* $?b$, *a state* s_2' *and a real number* ϵ' *such that* $s_1 \xrightarrow{(?b,\overline{x})}_{\epsilon'} s_2'$.

We say that a sequence

$$s_0 \xrightarrow{(a_1,\overline{x_1})}_{\epsilon_1} s_1 \xrightarrow{(a_2,\overline{x_2})}_{\epsilon_2} \cdots \xrightarrow{(a_n,\overline{x_n})}_{\epsilon_n} s_n$$

of consecutive transitions starting in the initial state of the automaton A *is a* $\triangle$-trace *of* A *if* $\epsilon = \triangle\{\epsilon_1, \ldots, \epsilon_n\}$ *is greater than zero and we write* $s_0 \xrightarrow{(a_1,\ldots,a_n,\overline{x_1}\ldots,\overline{x_n})}_\epsilon s_n$.

Example 2 *Consider the automata* Heart *given in Figures 1, 2 and 3 and the* 4*-tuple of fields: (gender, age, bpm, rr). The* 4*-tuple of values (*x_1, x_2, x_3, x_4*) indicates that the field gender is set to* x_1 *(we consider that* 0 *denotes males and* 1 *denotes females), age is set to* x_2, *the current bpm is set*

Action	Value	gender	age	bpm	rr
$?checkGender$	$(1,\cdot,\cdot,\cdot)$	1	0	0	0
$?checkAge$	$(\cdot,32,\cdot,\cdot)$	1	32	0	0
$!recBPM$	$(\cdot,\cdot,\cdot,\cdot)$	1	32	0	0
$?readBPM$	$(\cdot,\cdot,62,\cdot)$	1	32	62	0
$?readBPM$	$(\cdot,\cdot,69,\cdot)$	1	32	69	0
$!recRR$	$(\cdot,\cdot,\cdot,\cdot)$	1	32	69	0
$?readRR$	$(\cdot,\cdot,\cdot,977)$	1	32	69	977

Table II
CHANGES IN THE VALUES OF THE FIELDS

to x_3 heartbeats per minute and the current rr is set to x_4 milliseconds. We can observe a trace such as

$$(?checkGender,(1,\cdot,\cdot,\cdot)),(?checkAge,(\cdot,32,\cdot,\cdot)),$$
$$(!recBPM,(\cdot,\cdot,\cdot,\cdot)),(?readBPM,(\cdot,\cdot,62,\cdot)),$$
$$(?readBPM,(\cdot,\cdot,69,\cdot)),(!recRR,(\cdot,\cdot,\cdot,\cdot)),$$
$$(?readRR,(\cdot,\cdot,\cdot,977))$$

where inputs are preceded by $?$ and outputs are preceded by $!$. The value of each field is overwritten with each new element of the trace. The evolution of the values appears in Table III.

IV. CASE STUDY

In this section we present the application of our fuzzy automata in a real scenario: prediction of heart problems. We define the automaton *Heart* that is able to alert about the level of risk of a patient, according to the available information and physical evidence collected from his electrocardiograms (ECGs). The information managed by the automaton is:

- Gender. We have 2 groups: Man or Woman.
- Age. We have 8 groups of age: the ranges were previously introduced in Example 1 of Section III.
- Heartbeats. The range of correct heartbeats (beats per minute) for a healthy patient, according to his gender and age.
- RR waves. The range of correct RR waves duration (milliseconds) for a healthy patient, according to his gender, age and heartbeats.

Additionally, we consider that our set of actions consists of the following 7 operations:

- $?checkGender$: Reads the gender of the patient.
- $?checkAge$: Reads the age of the patient.
- $?readBPM$ and $?readRR$: Read the current heartbeats per minute and RR wave longitudes, respectively, of the patient.
- $!recBPM$ and $!recRR$: Indicate that the automaton is ready to receive information concerning beats per minute and RR waves duration, respectively.
- $!recordAlarm$: Record an alarm with the values of the four fields in that moment.

Our automaton has a total of 83 states and 146 transitions. Initially, our automaton has two transitions: one per gender (see Figure 1). Due to the limited space, we concentrate on the part of the automaton corresponding to male patients (see Figure 2) because the most relevant results of the study are related to subjects of this gender. There are 8 different transitions, one per age group. All the transitions lead to states having similar structure but with different fuzzy constraints to manage the data of the ECG (heartbeats and RR waves). In Figure 3 we depict the A_6 branch, that is, the part of the automaton corresponding to analyze the data of male patients in the range $60-69$ years old. We process BPMs until we find a potential problem during a one minute slot, having the samples a duration of 30 minutes. Therefore, the maximum number of problems per sample is equal to 30 (just when we have a problematic observation each minute of the sample). Then, we process RR waves until either we confirm the problem or we continue receiving BPMs. Let us remark that the transitions labelled by an output and departing from states q_5 and q_6 will be executed only when all the RR wave durations have been processed. Also, if we reach s_6 then we know that we have to raise an alarm but first we discard the pending RR wave durations.

Thanks to the data from other authors, we (manually) established the fuzzy constraints associated with transitions. Age and heartbeats data have been gathered from the work of Rijnbeek et al. [19]. In the case of the age, the δ value is obtained from the 20% of the highest value of each age range. As we previously mentioned, there are numerous factors that provoke the wrong classification of a patient according to his age. With this percentage, we extend the options to classify a patient according not only to his real age but also to other groups. This is so because a patient can present a healthier or more aging heart than a patient of his age group. Therefore, he needs to be analyzed with different parameters, as shown in Example 1. In the case of heartbeats, we had the median, 2nd percentile and 98th percentile from the database, but they were not applicable as limits for our automaton because they are not characteristic data of the sample of patients. For that reason, we applied the estimations made by Hozo et al. [33]: if the size of a sample exceeds 25 (the database that we use [19] has information from 13354 patients), the median itself is the best estimator for the mean and the best estimator for the standard deviation (SD) is

$$\sigma \approx \frac{b-a}{6}$$

where a is the smallest value of the sample (the 2nd percentile in our case) and b is the largest value of the sample (the 98th percentile in our case). Therefore, our limits are based on the median of each range and δ is based on the SDs of each range.

Concerning RR waves, we have used the data from the work of Haarmark et al. [20]. The problem in this case was that we only had the information of the RR waves duration for the patients of the age range $[30-39]$. So, if we would have used these limits for all the patients, then the prediction would have been erroneous. Therefore, we appealed to the work of Khachaturian et al. [34] which estimate that the duration of the RR waves can be derived from heartbeats (bpm). So, our limits are based on the application of the following formula

$$RR_{ms} \approx \frac{60000}{bpm}$$

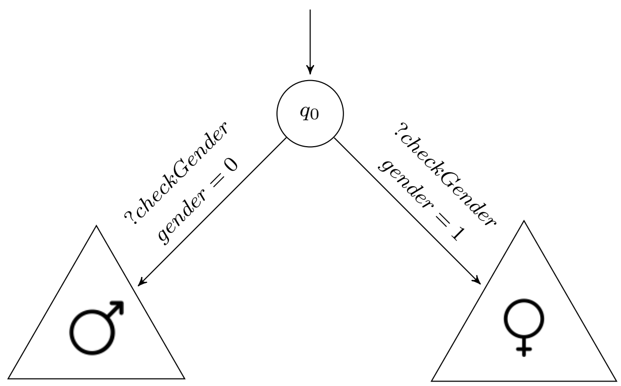

Figure 1. Fuzzy automaton *Heart*.

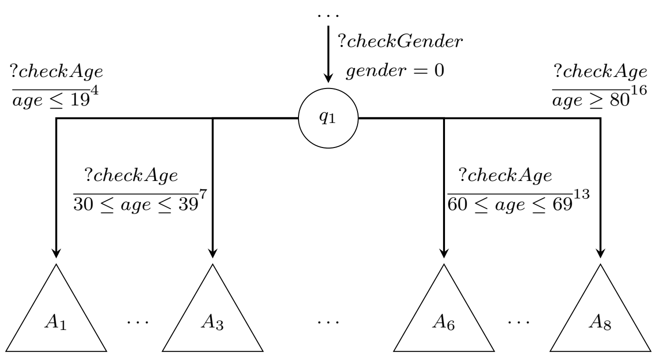

Figure 2. Male branch of the automaton.

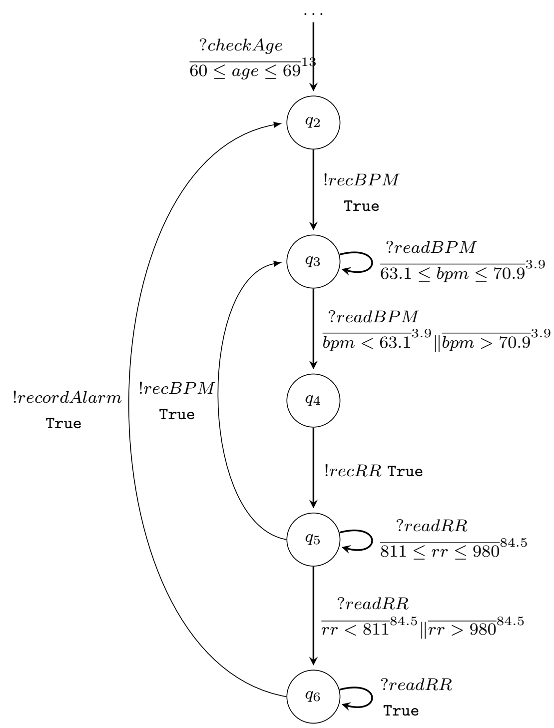

Figure 3. A_6 branch for patients between 60 and 69 years old.

to the bpm data obtained from the work of Rijnbeek et al. [19].

As previously mentioned, if at any moment of the study the state q_6 of any branch is reached, then the automaton will record the current state of the patient as a case in which he suffers the risk of having a serious heart problem.

In order to automatise the process of checking information against the automaton, we implemented a tool, in Java, to present summaries of patients from their ECGs. This application, the complete automata and some useful information about this study are available at https://bitbucket.org/azcama/heartdiagnosis.

In order to verify the applicability of our automaton, we used the "MIT/BIH Arrhythmia Database Directory": 48 ECG recordings with a duration of 30 minutes from the Massachusetts Institute of Technology - Beth Israel Hospital arrhythmia database [35]. All of them present some heart pathology but 48% of these samples have been annotated in the database as representative cases of routine clinical recordings while the remaining 52% reflect uncommon cases of arrhythmias. Due to space limitations, we focus again on patients between 60 and 69 years of age and show the results of all of the patients that have been studied according to the requirements of this branch. In Table IV we give the ID of the patient, his age, gender, GoC using the Gödel and Hamacher t-norms, the number of alarms detected and the number of points of interest (PoI) that are highlighted in the medical history of the patient by the healthcare professionals. We considered these points as an equivalence of alarms that are necessary to take into account during the study of the ECGs. The main difference between them is that the duration of the alarms does not exceed a minute, whereas the duration of a PoI can be longer than a minute. As a consequence, PoIs give a more general view of the ECG and alarms provide more accurate information about them. The reasons for using the Gödel and Hamacher t-norms are because the first help us to control the minimal GoC that a patient can obtain respect to the automaton, while the second t-norm is strongly recommended for aggregation operations [36]. Therefore, we would like to find the "average" of the most real GoC for a patient.

Note that the number of alerts detected by our automaton is higher than the number of PoIs diagnosed by the physicians. These PoIs are a brief summary of the most relevant aspects of the ECGs and their main goal is to help the interpretation of the data. However, they do not indicate the exact BPM or RR wave that provoke the highlighting of that point. This is an important improvement with respect to any previous work that we identified: besides detecting the same PoIs as the healthcare professionals, we give an extended summary with the heartbeats and RR waves that trigger the alert. This new information will be helpful for analyzing the arrhythmia pattern that can suffer the patient in order to decide the most suitable treatment that has to be provided to him.

There are a total of 31 samples than can be studied in this range. Analyzing the GoC obtained in the field *age* with this branch, we found that 14 of the patients obtain a GoC of 1, 7 of them obtain a GoC under 1 because they are under 60 years old

ID	Gender	Age	GoC - $\overline{\wedge}$	GoC - $\circledast$	Alarms	PoI
100	0	69	0.006	0.41	28	4
101	1	75	0.011	0.49	27	5
104	1	66	0.23	0.3	8	7
105	1	73	0.02	0.69	30	9
107	0	63	0.3	0.98	22	5
109	0	64	0.02	0.75	30	12
111	1	47	0.01	0.05	14	9
112	0	54	0.005	0.44	30	3
114	1	72	0.11	0.77	21	11
116	0	68	0.02	0.92	30	7
117	0	69	0.1	0.994	15	3
118	0	69	0.02	0.997	27	11
119	1	51	0.04	0.35	27	6
122	0	51	0.003	0.35	30	3
123	1	63	0.1	0.99	30	4
124	0	77	0.02	0.42	30	10
200	0	64	0.02	0.9	30	10
201	0	68	0.02	0.994	30	12
202	0	68	0.07	0.97	29	12
205	0	59	0.18	0.28	30	9
209	0	62	0.15	0.982	30	8
213	0	61	0.01	0.99	30	10
214	0	53	0.005	0.4	29	10
215	0	81	0.124	0.129	30	10
217	0	65	0.7	0.99	30	12
223	0	73	0.35	0.71	30	11
228	1	80	0.16	0.19	11	9
231	1	72	0.04	0.78	30	5
232	1	76	0.09	0.38	29	11
233	0	57	0.04	0.77	30	10
234	1	56	0.06	0.76	30	7

Table III
RESULTS OF THE PATIENTS STUDIED WITH BRANCH A_6.

and 10 of them obtain a GoC under 1 because they are older than 69. These 17 patients that do not belong to the considered age range obtain higher Gödel and Hamacher t-norm values in their correct age range. In the case of the other 14 patients, the ones between 60 and 69, there are 3 worth to mention cases: patients 100, 104 and 109. This information is depicted in Figure 4 where we present the highest GoC obtained for each patient according to his corresponding age branch. Following the results derived from the application of the Hamacher t-norm, most of the values obtain a GoC over 0.8, specifically the 73% of the patients. They are correctly classified according to the age branch where they belong. Notwithstanding this fact, there is a group of patients, the 27% of the sample, that would be better classified in another age branch because they present uncommon cases of arrhythmias in accordance with their age. This is the case of the patients 100, 104 and 109.

In the cases of patients 100 and 109 we can find a common pattern: males, between 60 and 69 years and with a Hamacher value below 0.8 (patient 100 with 0.41 and patient 109 with 0.75). If we check their results in the spreadsheet "patient-Data" of the repository, then we see that they obtain higher values in the age range of 70s. These could represent clear cases of premature aging of the heart tissue because their ECGs represent patterns of older people. Moreover, the values of the Gödel t-norm are quite low (under 0.01). This means that there are cases where these patients are classified as a 60s person but without any relevance. On the contrary, patient 104 reflects the opposite. Despite being a female patient between

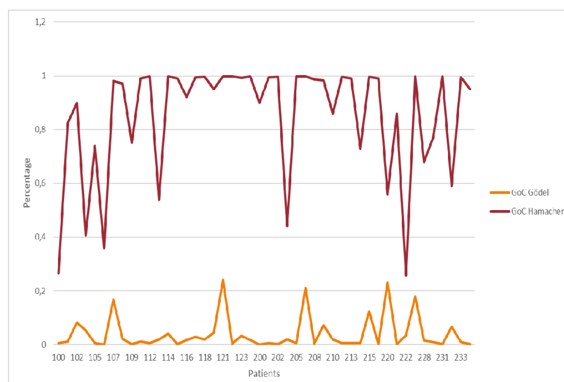

Figure 4. GoCs obtained from Gödel and Hamacher t-norms.

60 and 69 years with a Hamacher value under 0.8 (0.3), her highest GoC corresponds to a woman on her 50s (see data from the spreadsheet "patientData"). Also, her Gödel value is very similar to her Hamacher, that is, most of the times this woman has been classified as a 60s person, but without a consistent GoC. This could be the case of a sport female that regardless of having a healthy life style, suffers some heart pathologies corresponding to a younger person.

V. CONCLUSIONS

Healthcare is one of the most productive areas in research, but there are still plenty of pending issues to solve. In case of cardiology, many patients do not receive a correct diagnosis despite of the advances in treatments and prediction systems. However, with the help of other areas this lack could be improved. In this paper, we proposed the combination of ECG data, age and gender of a patient to detect heart problems with a fuzzy automaton. The obtained results are very promising for promoting the use of fuzzy constraints during the analysis of the state of a cardiology patient. In fact, we were able to correctly classify patients with the pathologies of individuals corresponding to other age group. Taking into account these first results, our next goal is to improve the automaton with the application of data mining techniques. We expect to obtain more accurate information about the behavior of the heart in each gender and age. In addition to this improvement, we need better feedback from cardio-healthcare professionals, to complete a list of the most relevant aspects that they consider during a diagnosis, so that we can improve our automaton and turn it into a more helpful tool for the physicians of this area.

Acknowledgements

Research partially supported by the projects DArDOS (TIN2015-65845-C3-1-R (MINECO/FEDER)) and SICOMORo-CM (S2013/ICE-3006). The first author is also supported by a *Universidad Complutense de Madrid – Santander Universidades* grant. Finally, we would like the anonymous reviewers for the careful reading and useful suggestions.

References

[1] R. V. Binder, B. Legeard, and A. Kramer, “Model-based testing: where does it stand?” *Communications of the ACM*, vol. 58, no. 2, pp. 52–56, 2015.

[2] A. R. Cavalli, T. Higashino, and M. Núñez, “A survey on formal active and passive testing with applications to the cloud,” *Annales of Telecommunications*, vol. 70, no. 3-4, pp. 85–93, 2015.

[3] C. Baier and J.-P. Katoen, *Principles of Model Checking*. MIT Press, 2008.

[4] L. Lamport, “Who builds a house without drawing blueprints?” *Communications of the ACM*, vol. 58, no. 4, pp. 38–41, 2015.

[5] S. L. Smith and S. Cagnoni, Eds., *Genetic and Evolutionary Computation: Medical Applications*. Wiley, 2010.

[6] A. Camacho, M. G. Merayo, and M. Núñez, “Collective intelligence and databases in eHealth: A survey,” *Journal of Intelligent & Fuzzy Systems (in press)*, 2017.

[7] P. Hunter, A. Pullan, and B. Smaill, “Modeling total heart function,” *Annual review of biomedical engineering*, vol. 5, no. 1, pp. 147–177, 2003.

[8] Z. Jiang, A.Connolly, and R. Mangharam, “Using the virtual heart model to validate the mode-switch pacemaker operation,” in *32nd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, EMBC’10*. IEEE Computer Society, 2010, pp. 6690–6693.

[9] D. Méry and N. Singh, “Formalization of heart models based on the conduction of electrical impulses and cellular automata,” in *1st Int. Symp. on Foundations of Health Informatics Engineering and Systems, FHIES’11, LNCS 7151*. Springer, 2012.

[10] T. Chen, M. Diciolla, M. Kwiatkowska, and A. Mereacre, “Quantitative verification of implantable cardiac pacemakers over hybrid heart models,” *Information and Computation*, vol. 236, pp. 87–101, 2014.

[11] W. G. Wee and K. S. Fu, “A formulation of fuzzy automata and its application as a model of learning systems,” *IEEE Transactions on Systems Science and Cybernetics*, vol. 5, no. 3, pp. 215–223, 1969.

[12] J. N. Mordeson and D. S. Malik, *Fuzzy automata and languages: theory and applications*. Chapman & Hall/CRC, 2002.

[13] M. Doostfatemeh and S. C. Kremer, “New directions in fuzzy automata,” *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 175–214, 2005.

[14] F. Crespo, A. d. Encina, and L. Llana, “Fuzzy-timed automata,” in *IFIP 30th Int. Conf. on Formal Techniques for Distributed Systems, FMOODS/FORTE’10, LNCS 6117*. Springer, 2010, pp. 140–154.

[15] Y. Cha, G. Lee, K. Klarich, and M. Grogan, “Premature ventricular contraction-induced cardiomyopathy a treatable condition,” *Circulation: Arrhythmia and Electrophysiology*, vol. 5, no. 1, pp. 229–236, 2012.

[16] J. Mackay, G. Mensah, S. Mendis, and K. Greenlund, *The atlas of heart disease and stroke*. World Health Organization, 2004.

[17] F. Wilson, A. Macleod, P. Barker, and F. Johnston, “The determination and the significance of the areas of the ventricular deflections of the electrocardiogram,” *American Heart Journal*, vol. 10, no. 1, pp. 46–61, 1934.

[18] A. v. Hof, A. Liem, M. d. Boer, and F. Zijlstra, “Clinical value of 12-lead electrocardiogram after successful reperfusion therapy for acute myocardial infarction. Zwolle Myocardial Infarction Study Group,” *The Lancet*, vol. 350, no. 9078, pp. 615–619, 1997.

[19] P. Rijnbeek, G. van Herpen, M. Bots, S. Man, N. Verweij, A. Hofman, H. Hillege, M. Numans, C. Swenne, J. Witteman *et al.*, “Normal values of the electrocardiogram for ages 16–90 years,” *Journal of electrocardiology*, vol. 47, no. 6, pp. 914–921, 2014.

[20] C. Haarmark, C. Graff, M. Andersen, T. Hardahl, J. Struijk, E. Toft, J. Xue, G. Rowlandson, P. Hansen, and J. Kanters, “Reference values of electrocardiogram repolarization variables in a healthy population,” *Journal of electrocardiology*, vol. 43, no. 1, pp. 31–39, 2010.

[21] A. Jervell and F. Lange-Nielsen, “Congenital deaf-mutism, functional heart disease with prolongation of the qt interval, and sudden death,” *American heart journal*, vol. 54, no. 1, pp. 59–68, 1957.

[22] M. Packer, “Lack of relation between ventricular arrhythmias and sudden death in patients with chronic heart failure,” *Circulation*, vol. 85, no. 1 Suppl, pp. I50–6, 1992.

[23] I. Christov and G. Bortolan, “Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks,” *Physiological Measurement*, vol. 25, no. 5, p. 1281, 2004.

[24] K. Srinivas, B. Rani, and A. Govrdhan, “Applications of data mining techniques in healthcare and prediction of heart attacks,” *International Journal on Computer Science and Engineering*, vol. 2, no. 2, pp. 250–255, 2010.

[25] G. Bortolan, I. Jekova, and I. Christov, “Comparison of four methods for premature ventricular contraction and normal beat clustering,” in *32nd Annual Scientific Proceedings of Computers in Cardiology*. IEEE Computer Society, 2005, pp. 921–924.

[26] A. Wojtasik, Z. Jaworski, W. Kuźmicz, A. Wielgus, A. Wałkanis, and D. Sarna, “Fuzzy logic controller for rate-adaptive heart pacemaker,” *Applied Soft Computing*, vol. 4, no. 3, pp. 259–270, 2004.

[27] G. Dorn, “The fuzzy logic of physiological cardiac hypertrophy,” *Hypertension*, vol. 49, no. 5, pp. 962–970, 2007.

[28] B. Giardine, C. Riemer, R. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, and J. Taylor, “Galaxy: a platform for interactive large-scale genome analysis,” *Genome research*, vol. 15, no. 10, pp. 1451–1455, 2005.

[29] C. O’Brien, A. Pollett, S. Gallinger, and J. Dick, “A human colon cancer cell capable of initiating tumour growth in immunodeficient mice,” *Nature*, vol. 445, no. 7123, pp. 106–110, 2007.

[30] C. Andrés, L. Llana, and M. Núñez, “Self-adaptive fuzzy-timed systems,” in *13th IEEE Congress on Evolutionary Computation, CEC’11*. IEEE Computer Society, 2011, pp. 115–122.

[31] R. Alur and D. Dill, “A theory of timed automata,” *Theoretical Computer Science*, vol. 126, pp. 183–235, 1994.

[32] K. Larsen, P. Pettersson, and W. Yi, “Uppaal in a nutshell,” *International Journal on Software Tools for Technology Transfer*, vol. 1, no. 1-2, pp. 134–152, 1997.

[33] S. Hozo, B. Djulbegovic, and I. Hozo, “Estimating the mean and variance from the median, range, and the size of a sample,” *BMC medical research methodology*, vol. 5, no. 1, p. 1, 2005.

[34] Z. Khachaturian, J. Kerr, R. Kruger, and J. Schachter, “A methodological note: Comparison between period and rate data in studies of cardiac function,” *Psychophysiology*, vol. 9, no. 5, pp. 539–545, 1972.

[35] J. Welch, P. Ford, R. Teplick, and R. Rubsamen, “The Massachusetts General Hospital-Marquette Foundation hemodynamic and electrocardiographic database–comprehensive collection of critical care waveforms,” *Clinical Monitoring*, vol. 7, no. 1, pp. 96–97, 1991.

[36] M. Oussalah, “On the use of Hamacher’s t-norms family for information aggregation,” *Information sciences*, vol. 153, pp. 107–154, 2003.