

Using Bayesian Networks to improve the Decision-Making Process in Public Health Systems

Germanno Teles¹, Carina Oliveira², Reinaldo Braga², Luiz Andrade³,

Ronaldo Ramos², Paulo Cunha⁴, Mauro Oliveira²

¹State University of Ceará (UECE) - Fortaleza, Brazil

²Federal Institute of Ceará (IFCE) - Fortaleza, Brazil

³Federal University of Ceará (UFC) - Fortaleza, Brazil

⁴Federal University of Pernambuco (UFPE) - Recife, Brazil

Email: germanno@bnb.gov.br, {carina.oliveira,reinaldo.braga,ronaldo,mauro}@ifce.edu.br, odorico@saude.gov.br, prfc@cin.ufpe.br

Abstract - This paper proposes the use of Bayesian networks to support the decision-making process in public health systems. In particular, this paper presents LARIISA_Bay, a new component based on Bayesian networks that works together with LARIISA, a context-aware platform to support applications in public health systems. The main goal of the proposed component is to assist teams of health specialists in order to better diagnose diseases through data collected from users of LARIISA. As a case study, we focus on scenarios of dengue fever disease. We classify dengue cases into one of the following levels: normal, grave or emergency. Based on this classification, teams of health specialists can accurately make decisions, for example, to alert a health care agent to visit locations with a high incidence of the disease, to send a team of health specialists when a dengue emergency case has occurred, as well as give technical instructions on how to deal with specific cases. We present a prototype of LARIISA_Bay as well as the corresponding interfaces to support the interactions with the component. We compare the obtained results with real diagnosis of general practitioners. The results presented show the efficiency of the proposed approach.

Keywords—Bayesian Network, Public Health Systems, Dengue Fever Disease

I. INTRODUCTION

With the increasing advances in medical and computer sciences, it is noticed that the union of these two areas becomes indispensable for the development of efficient systems focused on the care and treatment of patients [1]. According to Sigulem *et al.* [1], since the beginning of computing there is a significant enthusiasm regarding the use of computers as an important tool to support medical diagnosis.

We add to this fact the recent challenges faced by many organizations, especially in the public sphere, to systematize the large amount and variety of data produced separately and then to generate useful information in a timely manner. In public health systems, this issue is even more challenging given the high complexity of the relation between different health systems, several diseases and various levels of medical care.

Considering that, several medical diagnosis solutions have been proposed. However, most of these solutions present a simplistic structure composed of a trivial “*disease x symptom*”

relation. For this reason, inference engines implemented in these solutions do not have a high computational complexity.

In this context, we present LARIISA_Bay as a new component of the LARIISA framework [3]. LARIISA_Bay is capable to improve public health systems in the diagnosis of medical diseases. To achieve this goal, it uses Bayesian networks, which are a powerful method for the construction of systems that rely on probabilistic knowledge. An example is the use of probabilistic networks in uncertainty modeling for medical diagnosis such as epidemiological diseases, heart disease, Alzheimer's, etc.

As a case study, we focus on scenarios of dengue fever disease. For instance, LARIISA_Bay can act for different aims: i) an important mechanism to accurately identify patients who must be treated immediately (i.e., patients with dengue hemorrhagic fever), ii) direct patients to specific hospitals for further treatment and analysis, iii) treatment of patients at home to avoid the overcrowding of hospitals, iv) identification of risk areas in order to adopt proactive preventions, etc. In other words, LARIISA_Bay increases the diagnosis efficiency of medical diseases and, consequently, optimizes the public health system as a whole.

To make the solution as realistic as possible, we have done several meetings with various health specialists to better understand the pathology specifics. These meetings facilitate the understanding of the procedures to diagnose the disease and the main factors that contribute to the worsening of the dengue fever disease.

Finally, we present a prototype of LARIISA_Bay and its corresponding evaluation. To achieve this goal, we have also implemented an interface to allow the interaction of three main actors in the health system: the patient, the health agent and the medical specialist. The collected data are used to feed the Bayesian network. Our results demonstrate the proposed solution efficiency in the classification of dengue cases into three main levels: normal (i.e., absence of the disease), grave (i.e., dengue fever) or emergency (i.e., dengue hemorrhagic fever). The results are compared to real diagnoses of general practitioners.

The remainder of this paper is organized as follows: We discuss the related work in Section II. Section III presents the LARIISA framework. Section IV describes in more details the LARIISA_Bay component. Section V presents the network modeling and our evaluation. Finally, Section VI concludes the paper.

II. RELATED WORK

In this section, we present the related work that makes use of Bayesian networks in the health system context. They demonstrate the relevance of using computational intelligence in the treatment of uncertainty in healthcare diagnosis.

A. Bayesian Agent for Surveillance of Hospital Infection - SAVIH [4]

SAVIH uses a Bayesian network to support the hospital manager to evaluate the risk of a patient to contract a hospital infection. Among other features, SAVIH presents the infection scenarios in different inpatient units allowing the prediction of risks based on the patient's disease and epidemiological characteristic. Besides, it enables the retrieval of past cases that are similar to new patients entering in the inpatient unit.

SAVIH was developed using the modeling shell Netica [5], which is a software package to solve problems using Bayesian networks.

B. System to Support Differential Diagnosis of Cephalgia [6]

This work presents a medical expert system that provides support to general practitioners, emergency physicians or residents in the differential diagnosis for cephalgia (i.e., headaches). As SAVIH, this system uses the Bayesian network approach and was developed using the modeling shell Netica. The base of the knowledge was build considering the classification criteria of the International Headache Society (IHS), taking into consideration the patient's symptoms and the estimated values of probabilities, provided by medical experts involved in the project.

The evaluation was performed through a comparison between the responses of medical experts and those obtained from the proposed system. For this evaluation, a set of medical patient records was randomly selected by the specialists involved in this project. The experimental results indicated that the proposed system was able to provide the same diagnosis of experts in 95% of the cases. When the same clinical cases were evaluated by general practitioners, 53% of the cases were correct. These results demonstrated that the developed system presented a good performance in the support of differential diagnosis of cephalgia.

III. LARIISA FRAMEWORK

In this section, we present LARIISA, a context-aware framework that provides intelligence governance for public health systems [3]. Based on Dey's definition of context [10], we consider *health context* as any information that can be used to characterize the situation of an entity in a health system. Thus, the main idea of LARIISA is to infer the data collected from users (e.g., patients, doctors, etc), such as those of medical patient records.

To assist these users in their daily tasks, context-aware applications can use elements of ubiquitous systems to obtain user context information. For instance, through the use of embedded sensors in mobile devices, the system can acquire spatio-temporal information of a user.

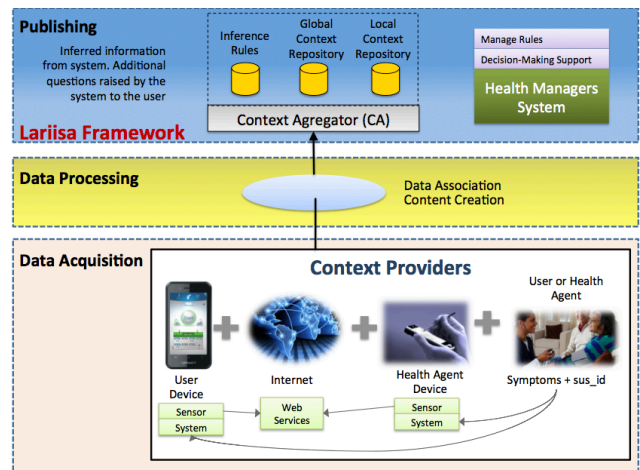


Fig. 1. Overview of the LARIISA Framework [9].

Figure 1 illustrates an overview of LARIISA, which is divided in three main layers. The first layer is called *Data Acquisition*. This layer concerns the information acquired from context providers, such as sensed data and medical patient records. For example, when a health agent is accessing the system by a mobile device from the patient's home, he can send information about the patient's health, detailing specific symptoms required by the system. An interesting observation is that patients can be identified in the system through a personal ID, such as the SUS_ID of the Brazil's Unified Health System [8]. After the data acquisition, the data is processed in the *Data Processing* layer. In this case, the system uses the raw data in order to capture all necessary diagnosis data [9]. Finally, after the acquisition and processing of data, the *Publishing* layer produces new contents and stores in the database of LARIISA.

In summary, the data inference in LARIISA is beneficial in the decision-making process in public health systems. For example, managers of public health systems (i.e., governors, mayors, hospital directors, doctors, health agents, etc.) can be aware of real-time risk situations (e.g., epidemiological emergencies) and adapt the procedures in a timely manner. Thus, five domains of governance can be achieved by LARIISA [3]:

- (1) Knowledge Management;
- (2) Systemic Normative;
- (3) Clinical and Epidemiological;
- (4) Administrative and
- (5) Shared Management.

IV. LARIISA_BAY COMPONENT

LARIISA_Bay is a new component based on Bayesian networks that works together with LARIISA. This new component is concerned with the treatment of uncertainty in the health system. Therefore, this work is related to the representation of context-sensitive information (i.e., collected data) as well as to the knowledge of experts, representing the ontology for LARIISA [10]. As a result, this component is able to assist team of specialists to better diagnose diseases according to the data collected from different users of the system. We highlight that the treatment of uncertainty in the data acquisition and the representation of knowledge have not yet been addressed by LARIISA.

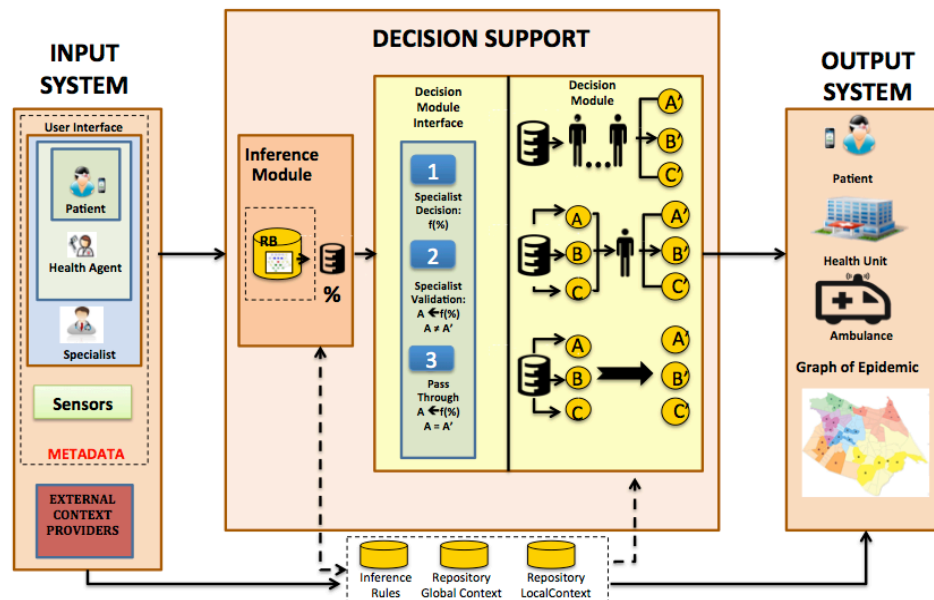


Fig. 2. The LARIISA_Bay Phases.

Figure 2 illustrates the proposed phases of LARIISA_Bay. To better illustrate these phases, we describe examples of the dengue fever disease, which corresponds to the case study of this work.

- **The Input System:** it corresponds to an information gathering interface that allows the interaction of three different decision makers: the patient (i.e. user of the system), the health agent and the specialist. In Figure 2, note that the specialist interface contains the health agent interface, which in turn contains the patient interface. In this initial phase, sensor can also be used, for example, to monitor patients' vital signs. As a result, metadata is created from the gathered information. It is also possible to use external context providers.
- **Decision Support:** the Inference Module corresponds to the Bayesian network modeled to support the decision making of medical teams. The details of the network modeling are presented in Section V. Basically, the Inference Module has two main purposes in the context of our case study:

(1) To support the diagnosis of the medical staff, filtering probable cases of dengue in three levels of classification, which are:

- Normal, for patients without dengue fever;
- Grave, for patients with dengue fever disease;
- Emergency, for patients with dengue hemorrhagic fever.

(2) To support the diagnosis of dengue fever outbreaks/epidemics in specific regions (i.e., risk areas).

Next, we have the Decision Module Interface, which offers three different application scenarios:

(1) *Specialist Decision* - it considers the existence of a team of experts able to better diagnose the dengue fever disease according to the received information. Based on the result of the Inference Module, the team of experts can take the most appropriate decision in relation to a particular patient;

(2) *Specialist Validation* - in this scenario, the result of the Inference Module is filtered/validated by an specialist, instead of analyzed, as in the first scenario;

(3) *"Pass Thought"* - in this scenario, one of the following decisions can be made: to leave the decision making to the system or to wait for a health specialist to take the decision.

- **The Output System:** it corresponds to the procedures that can be made after the Decision Support phase in order to optimize the public health system as a whole. We can mention the following procedures as examples:

- To send guiding procedures to the patient;
- To send a health agent to the patient's home;
- To send an emergency health team to the patient's home;
- To identify risk areas (graph of the epidemic) in order to adopt proactive preventions.

V. NETWORK MODELING

In this section, we introduce the details necessary to understand the network modeling. Then, we present the evolution and quantification of the Bayesian network.

A. Bayesian Networks

In this section, the formalism of Bayesian networks is introduced. Furthermore, the basic methods for their construction are reviewed.

Bayesian Networks (BN) are directed and acyclic graphs that allow the representation of the joint distribution of probabilities for a set of random variables. This paper focuses on BN with discrete variables. BN with discrete variables satisfy the Markov condition [10], which states that any node in a Bayesian network, given its parents, is conditionally independent of its non-descendants.

An important aspect of a BN is its structure (topology of the graph), which enables the representation of complex relationships between graphically and intuitively variables. The

graphical structure of a BN facilitates the understanding of relationships between variables in its domain. It also allows the combined use of information obtained from expert knowledge with historical data to obtain the joint probability distribution of the network.

BN used in data sorting problems are called Bayesian classifiers. Basically, if we have Bayesian classifiers with discrete variables (e.g., $\{A_1, A_2, \dots, A_n, C\}$), then one of them is the variable class (e.g., C , the response variable) and the others are the attributes (e.g., $\{A_1, A_2, \dots, A_n\}$, the predictor variables).

The excessive number of probabilities required for the quantification of a network is one of the greatest challenges in the practical application of Bayesian networks. For instance, in a node X with k categories, the number of probabilities to be specified is

$$\text{Numbers of elements of the table}(X) = P(x_i) \prod_{pa(X)} n_{category\ pa(X)} \quad (1)$$

However, for certain types of nodes, these probabilities can be calculated from other instead of being directly specified. The *noisy-or* model [12] allows such a calculation, with the restriction that the parents of the node may independently contribute to the probability. Therefore, the combination of several parents contributes cumulatively on the probability of the node.

The noisy-or model allows the calculation of the joint table of conditional probabilities from the probability $P(D|R_i)$ for each parent node R_i , such as R_1, R_2, \dots, R_n are the possible causes of a disease of a node (D).

In the case of having only binary nodes R_i , with stating r_i (true) and \bar{r}_i (false), these conditional probabilities are known as sensitivity $P(d|r_i)$ and specificity $P(\bar{d}|\bar{r}_i)$ that often are available in public studies. However, the joint probabilities $P(D|R_1, R_2, \dots, R_n)$ are more difficult to obtain from experts or bibliographic information because they involve a high number of conditional combinations.

The expression that represents the probability of each factor causing the disease is calculated as:

$$p_i = P(d|r_i \text{ alone}) = P(d|\bar{r}_1, \bar{r}_2, \dots, r_i, \dots, \bar{r}_n) \quad (2)$$

$$P(d|H) = 1 - \left(\prod_{R_i \in H^+} [1 - p_i] \right) \quad (3)$$

Where H is a configuration of (R_1, R_2, \dots, R_n) and H^+ is set as true.

B. Case study: Dengue Fever Disease

As a first step in the network modeling process, we have obtained the structural model of the network, using the knowledge of physicians as well as bibliographic sources in the subject of dengue fever disease.

To accurately model the network, it is important to identify which information from the medical point of view has to be represented as a variable in the network. Besides, it is fundamental to identify the causal relationships between these variables. To perform these objectives, the information is placed in an initial network structure. Then, an iterative refinement is made in order to obtain the final network structure for quantification. We highlight that, to achieve this final network structure, it is necessary the insertion and removal of nodes and arcs and other modifications.

According to our research made with physicians and bibliographic sources, a patient is suspected of having contracted the dengue fever disease when he/she presents acute

febrile illness with a maximum incubation period of seven days, accompanied by at least two of the following symptoms: headaches, retro-orbital pain, myalgia, arthralgia, prostration or rash. These symptoms can be associated or not with bleeding or bleeding with positive epidemiological history. Another indication is the patient has been in a dengue transmission area in the last fifteen days or in the presence of *Aedes aegypti* mosquito [13].

Other pathological symptoms observed are severe and continuous abdominal pain. The patient does not support superficial abdominal palpation and may prevent movements of sitting and walking. Besides, the patient presents orthostatic hypotension and/or syncope, persistent vomiting and does not the ingestion of any substance, including water. Thus, the patient presents decreased diuresis due to the dehydration caused by vomiting and/or hypotension. The patient has an enlarged liver (i.e., painful hepatomegaly) and has bleeding in respiratory mucosa and/or digestive system. The patient has also difficulty in maintaining alertness and presents psychomotor agitation. It presents hypothermia. It presents a sudden increase in hematocrit (proportional imbalance of water and blood cells) and abrupt platelet decrease. Finally, the patient presents respiratory distress, in which the patient has difficulty breathing, which may be due to bleeding or lung edema.

The occurrence of dengue is directly related to some risk factors, such as places without adequate sanitation and precarious garbage collection.

The accurate diagnosis of dengue is carried by serology. However, this test is performed in laboratories and hospitals. For this reason, the rapid diagnosis is important in order to referring patients for treatment. The main tests are serology and blood count.

C. Network Modeling Evolution

In this work, the network construction was done in two main steps in an iterative manner.

In the beginning, the model was restricted to the dengue fever pathologies. The main purpose of this initial model was to pre-select the network nodes and identify which nodes could indicate another symptom.

The next step in the network modeling was to identify and classify patients with symptoms of *dengue fever* or *dengue hemorrhagic fever*. In this step, the nodes in the network (the dengue symptoms) were classified according to three main profiles: patient, health agent and specialist. At this point, an interesting observation is that some symptoms could indicate the severity of the dengue fever disease.

In relation to the categories, two of them are defined for all nodes: *Yes* and *No*. Such representation is appropriate for diseases with several factors of cause. For example, Fig. 3 presents the final modeled network. Note that for the node *Dengue_hemorrhagic* the symptoms are: *abrupt_platelet_decrease*, *Sudden_increase_of_hematocrit*, *Hypothermia*, *Respiratory_Distress*, *Decreased_diuresis*, *irritability_and_drowsiness*, *Painful_hepatomegaly*, *Persistent_Vomiting*, *Orthostatic_hypotension*, *poor_garbage_collection_system_places*, *Severe_and_continuous_abdominal_pain*, *places_without_adequate_sanitation* and *Bleeding_from_mucosa*. In this paper, the noisy-or model was used for the nodes *Dengue_hemorrhagic* and *Dengue_fever*.

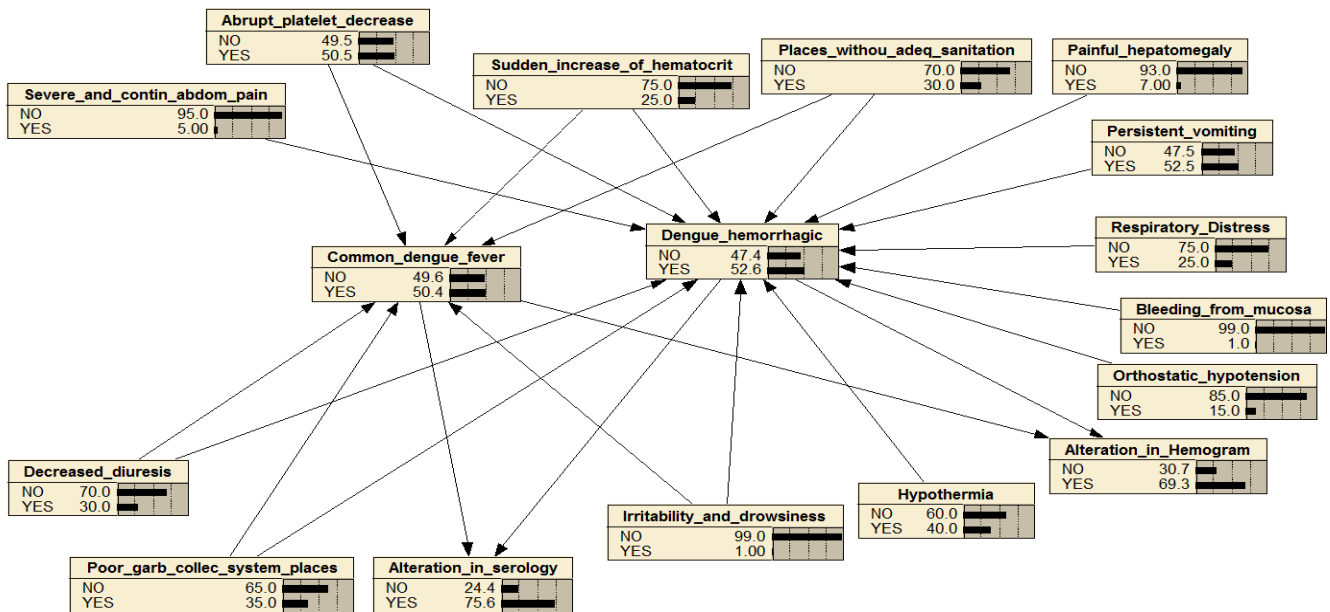


Fig. 3. Final model of the network created by Netica.

All things considered, Fig. 3 presents the final modeled network, which is used as the basis of inferences in LARIISA_Bay.

D. Probabilities

For the quantification of the Bayesian network proposed in this paper, it is used the direct description of probabilities presented in the following tables filled by expert physicians based on their professional experience and also on bibliographic sources. This method enables the fast computation of probabilities. To increase the coherency of values, some probability values are corrected during the evaluation.

Only three nodes have conditional probability related respectively to the parents in the LARIISA_Bay network. To specify these probabilities, the nodes are divided in two groups: diagnostic tests and pathologic symptoms. The nodes of the pathologic symptoms group represent the characteristics of the disease that, even though they are difficult to measure, they are indispensable to do the modeling.

TABLE I. PRIOR PROBABILITY

Node	Yes	No
Abrupt_platelet_decrease	1%	99%
Painful_hepatomegaly	7%	93%
Dor_adominal_intensa_continua	5%	95%
Decreased_diuresis	30%	70%
Orthostatic_hypotension	15%	85%
Irritability_and_drowsiness	1%	99%
Hypothermia	40%	60%
Poor_garbage_collection_system_places	35%	65%
Respiratory_Distress	25%	75%
Sudden_increase_of_hematocrit	25%	75%
Persistent_Vomiting	5%	95%

TABLE II. CONDITIONAL PROBABILITY

Variable Condition Categories	Blood_Count_Change	
	Yes	No
Dengue = Yes	87%	13%
Dengue = No	62%	38%

TABLE III. CONDITIONAL PROBABILITY

Variable Condition Categories	Serology_Change	
	Yes	No
Dengue = Yes	99%	1%
Dengue = No	9%	91%

E. Evaluation

For the evaluation of our approach we have considered the network generation based on two types of acquired information: the patient informing the symptoms; and the health specialist adding new technical data besides the patient information.

Firstly, the network evaluation is done with the information provided only by the patient, in three different cases: Low Risk, Medium Risk and High Risk. We have obtained the results showed in Table IV. The line called *Network* represents the result provided by the network for the node Dengue. The line called *General Practitioner* is the probability attributed by the general practitioner in relation to the same node. Finally, the line "Do you Agree?" shows the opinion of the general practitioner in relation to the network accuracy. This first result shows that the network percentage approximates the general practitioner percentage, which confirms the efficiency of our approach.

The second network evaluation took into account the network evaluation based on the data provided by a health agent. In this case, we have obtained the results showed in Table V. In this case, the results are also favorable to our approach.

TABLE IV. COMPARISON OF THE NETWORK AND GENERAL PRACTITIONER USING PATIENT INFORMATION

	Low Risk	Medium Risk	High Risk
Network	9.89%	24.8%	92.2%
General practitioner	5%	23%	86%
Do you agree?	Partially	Yes	Yes

TABLE V. COMPARISON OF THE NETWORK AND GENERAL PRACTITIONER USING PATIENT AND HEALTH AGENT INFORMATION

	Low Risk	Medium Risk	High Risk
Network	1.63%	11.1%	99.7%
General practitioner	5%	27%	96%
Do you agree?	Partially	Yes	Yes

VI. CONCLUSIONS

The decision-making process on health system becomes a complex process because it involves many variable surrounded by uncertainties. These are both inherent to the various actors in the system (user, health agent, physicians, hospital managers, secretaries, etc) as marginal factors to these actors (seasonal epidemics, economic and social aspects).

The contribution of this paper is regarding how the information uncertainties are treated. Therefore, it relates to both key-concepts of LARIISA: the representation context-aware based system and representation of the knowledge by ontology. It demonstrated, with the introduction of LARIISA_Bay, a new component of the inference mechanism of the LARIISA project.

Three approaches should be highlighted on the concept and development of LARIISA_Bay:

- This component assists a team of specialists to better diagnoses dengue cases as it collects data from system users, classifying them as: normal (absence of the disease), grave (presence of dengue fever) and emergency (dengue hemorrhagic fever). From this classification, team of experts can make the decision more accurately. For example, the immediate action of sending an ambulance or to ask a health agent to visit the patient or, simply, giving instructions on how to deal with the case, accordingly to the classification.
- The creation of the structure of the Bayesian network and the quantification of probabilities with the support of experts. They were obtained based on bibliographic data and through meetings with physicians. The probabilities were obtained through technical clarification, questionnaires and subsequent correction of initial values.
- The proposed model involves starting from mobile interfaces to the placement of metadata that are able to feed the table of probabilities to Bayesian network, created from consults to various expert professionals of dengue fever worsening. Three interfaces developed (user, health agent and expert) are related to three different metadata that translate the different levels of access to information and decision ability. That is, even though the final decision (dengue fever diagnosed) is

the expert's, the health agent and the user are also decision-makers in restricted cases.

Finally, it is important to highlight that the same solution proposed in the paper could be analyzed and evaluated taking as reference an ontological domain model built to meet the strategic goals of the study area. The research of hybrid mechanisms, using Bayesian networks and ontology, makes the prediction more refined and enables indirect inferences that are difficult to obtain without a model based on ontology. In this case, the Bayesian network result could feed the ontological model, e.g. based on this database, it generates the resulting inference to the LARIISA ontology using semantics research strategies in the base information of modeling with ontology.

REFERENCES

- [1] L. S. Petersen, P. Bertelsen, C. Bjornes, "Cooperation and communication challenges in small-scale eHealth development projects" in International Journal of Medical Informatics, vol. 82, no. 12, 2013.
- [2] D. Sigulem, M. S. Anção, M. P. Ramos, B. F. Leão, "Sistema de apoio à decisão em medicina", 1998. Available in: http://www.virtual.epm.br/material/tis/curr-med/sad_html/sistema.htm. Accessed in 01/04/2014.
- [3] M. Oliveira, O. M. Andrade, C. G. Hairon, R. C. Moura, S. Fernandes, J. Bringel, J. Gensel, H. Martin, C. Sicotte, J. L. Denis, "A context-aware framework for health care governance decision-making systems: A model based on the Brazilian digital TV" in IEEE IREHSS/WoWMoM, 2010.
- [4] M. Venicius, "Agente bayesiano de apoio à vigilância de infecção hospitalar – SAVIH". Master's Thesis, Department of Computer Science, UFSC, 2002.
- [5] Netica Application, Norsys Software Corporation. Available in <http://www.norsys.com/netica.html>. Accessed in 01/04/2014.
- [6] R. V. Toledo, "Sistema de apoio ao diagnóstico diferencial de cefaléia". Master's Thesis, Department of Computer Science, UFSC, 2000.
- [7] A. K. Dey, G. D. Abowd, "Towards a better understanding of context and context-awareness" in Workshop on the What, Who, Where, When and How of Context-Awareness, 2000.
- [8] SUS - Sistema Único de Saúde do Brasil, Available in: <http://portal.saude.gov.br>. Accessed in 01/04/2014.
- [9] L. M. Gardini, R. B. Braga, J. Bringel, R. M. C. Andrade, C. T. Oliveira, H. Martin, O. M. Andrade, M. Oliveira, "Clariisa, a context-aware framework based on geolocation for a health care governance system" in IEEE HealthCom, 2013.
- [10] M. Oliveira, O. M. Andrade, F. Antunes, L. M. Gardini, "Applying Ontology And Context Awareness Concepts On Health Management System: A Dengue Crisis Study Case" in IADIS, 2013.
- [11] R. E. Neapolitan, "Learning Bayesian Networks", New Jersey, 2004.
- [12] D. Heckerman, J. Breese, "A new look at causal independence" in UAI, 1994.
- [13] Brazilian Ministry of Health. "Programa Nacional de Controle da Dengue". Available in: <http://portal.saude.gov.br>. Accessed in 01/04/2014.