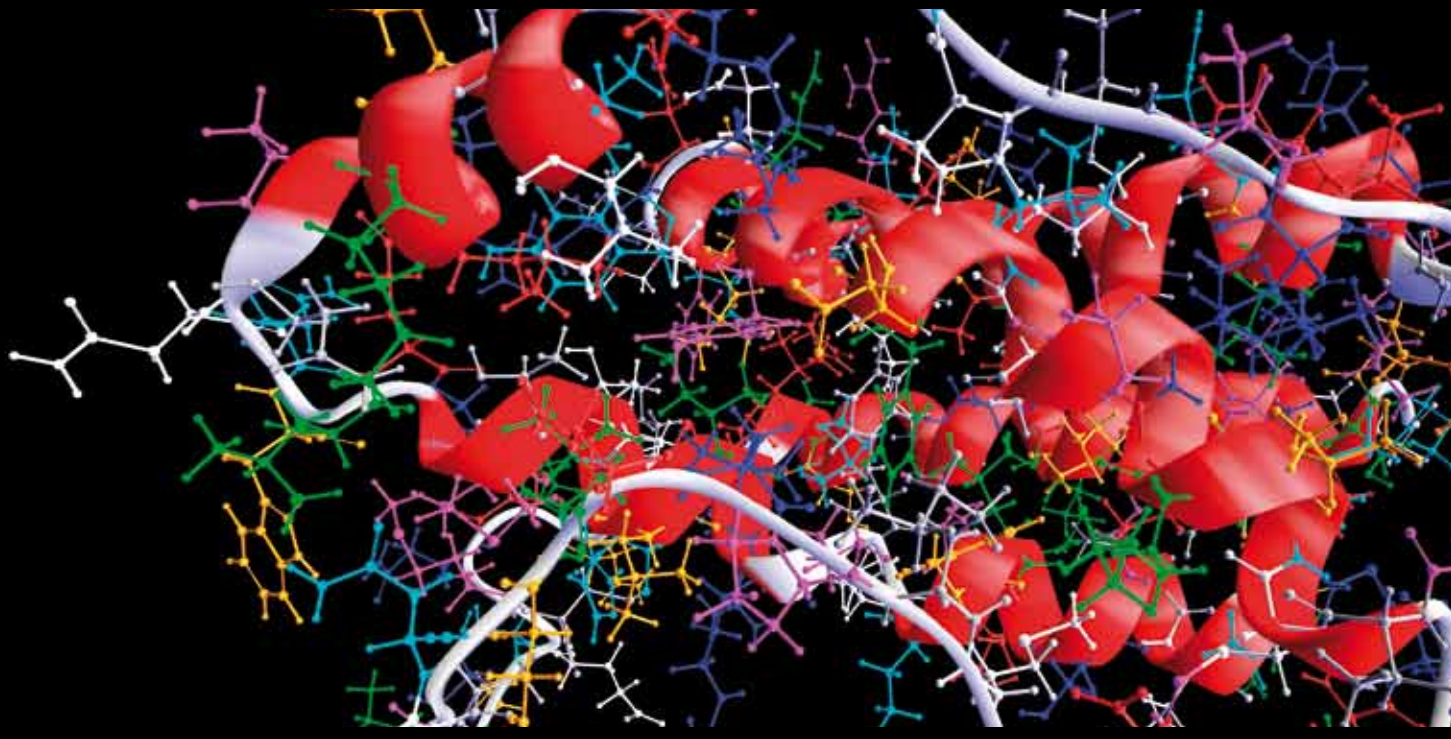
Computational and Mathematical Methods in Medicine

Methods and Models for Diagnosis and Prognosis in Medical Systems

Guest Editors: Angel García-Crespo, Giner Alor-Hernández, Linamara Battistella, and Alejandro Rodríguez-González



Methods and Models for Diagnosis and Prognosis in Medical Systems

Computational and Mathematical Methods in Medicine

Methods and Models for Diagnosis and Prognosis in Medical Systems

Guest Editors: Angel García-Crespo, Giner Alor-Hernández, Linamara Battistella, and Alejandro Rodríguez-González

Editorial Board

Contents

Methods and Models for Diagnosis and Prognosis in Medical Systems, Angel García-Crespo, Giner Alor-Hernández, Linamara Battistella, and Alejandro Rodríguez-González
Volume 2013, Article ID 184257, 2 pages

Evaluation of the Diagnostic Power of Thermography in Breast Cancer Using Bayesian Network Classifiers, Cruz-Ramírez Nicandro, Mezura-Montes Efrén, Ameca-Alducin María Yaneli, Martín-Del-Campo-Mena Enrique, Acosta-Mesa Héctor Gabriel, Pérez-Castro Nancy, Guerra-Hernández Alejandro, Hoyos-Rivera Guillermo de Jesús, and Barrientos-Martínez Rocío Erandi
Volume 2013, Article ID 264246, 10 pages

Design and Customization of Telemedicine Systems, Claudia I. Martínez-Alcalá, Mirna Muñoz, and Josep Monguet-Fierro
Volume 2013, Article ID 618025, 16 pages

Mobile Personal Health System for Ambulatory Blood Pressure Monitoring, Luis J. Mena, Vanessa G. Felix, Rodolfo Ostos, Jesus A. Gonzalez, Armando Cervantes, Armando Ochoa, Carlos Ruiz, Roberto Ramos, and Gladys E. Maestre
Volume 2013, Article ID 598196, 13 pages

Statistical Evaluation of a Fully Automated Mammographic Breast Density Algorithm, Mohamed Abdolell, Kaitlyn Tsuruda, Gerry Schaller, and Judy Caines
Volume 2013, Article ID 651091, 6 pages

Development of an Expert System as a Diagnostic Support of Cervical Cancer in Atypical Glandular Cells, Based on Fuzzy Logics and Image Interpretation, Karem R. Domínguez Hernández, Alberto A. Aguilar Lasserre, Rubén Posada Gómez, José A. Palet Guzmán, and Blanca E. González Sánchez
Volume 2013, Article ID 796387, 17 pages

Detection of Structural Changes in Tachogram Series for the Diagnosis of Atrial Fibrillation Events, Francesca Ieva, Anna Maria Paganoni, and Paolo Zanini
Volume 2013, Article ID 373401, 11 pages

Normality Index of Ventricular Contraction Based on a Statistical Model from FADS, Luis Jiménez-Ángeles, Raquel Valdés-Cristerna, Enrique Vallejo, David Bialostozky, and Verónica Medina-Bañuelos
Volume 2013, Article ID 617604, 12 pages

Study of the Effect of Breast Tissue Density on Detection of Masses in Mammograms, A. García-Manso, C. J. García-Orellana, H. M. González-Velasco, R. Gallardo-Caballero, and M. Macías-Macías
Volume 2013, Article ID 213794, 10 pages

An Intelligent System Approach for Asthma Prediction in Symptomatic Preschool Children, E. Chatzimichail, E. Paraskakis, M. Sitzimi, and A. Rigas
Volume 2013, Article ID 240182, 6 pages

Ontology-Oriented Diagnostic System for Traditional Chinese Medicine Based on Relation Refinement, Peiqin Gu, Huajun Chen, and Tong Yu
Volume 2013, Article ID 317803, 11 pages

The iOSC3 System: Using Ontologies and SWRL Rules for Intelligent Supervision and Care of Patients with Acute Cardiac Disorders, Marcos Martínez-Romero, José M. Vázquez-Naya, Javier Pereira, Miguel Pereira, Alejandro Pazos, and Gerardo Baños
Volume 2013, Article ID 650671, 13 pages

Locomotor Development Prediction Based on Statistical Model Parameters Identification, A. V. Wildemann, A. A. Tashkinov, and V. A. Bronnikov
Volume 2012, Article ID 548208, 5 pages

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 184257, 2 pages
http://dx.doi.org/10.1155/2013/184257

Editorial

Methods and Models for Diagnosis and Prognosis in Medical Systems

Angel García-Crespo,[1] Giner Alor-Hernandez,[2] Linamara Battistella,[3] and Alejandro Rodríguez-González[4]

[1] *Computer Science Department, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés, Spain*
[2] *Division of Research and Postgraduate Studies, Instituto Tecnológico de Orizaba, Avenida Oriente 9, 852 Colonia Emiliano Zapata, CP 94320 Orizaba, VER, Mexico*
[3] *Medical School, Universidade de São Paulo, Avenida Dr. Arnaldo, 455 Cerqueira César, 01246903 São Paulo, SP, Brazil*
[4] *Bioinformatics at Centre for Plant Biotechnology and Genomics UPM-INIA, Pozuelo de Alarcón, Parque Científico y Tecnológico, UPM, Campus de Montegancedo, Carretera M-40, km 38, 28223 Madrid, Spain*

Correspondence should be addressed to Angel García-Crespo; angel.garcia@uc3m.es

Received 1 August 2013; Accepted 1 August 2013

1. Introduction

During years, the development of medical systems (as part of Clinical Decision Support Systems (CDSS)) has been one of the main research fields in biomedical informatics area. Different systems have been proposed to improve the quality of medical practice in several different areas such as pharmacy management, electronic health records, diagnosis systems, telemedicine, and medical imaging among others.

Several approaches and models have been developed since, the earliest 60s. Also several methods and techniques from the fields of artificial intelligence, decision theory, and statistics have been introduced into models of the medical management of patients (diagnosis, treatment, and follow-up); in some of these models, assessment of the expected prognosis constitutes an integral part. Typically, recent prognostic methods rely on explicit (patho)physiological models, which may be combined with traditional models of life expectancy.

The main goals of this special issue were the publication of new algorithms and mathematical methods, decision theories, techniques, and models based on probabilistic and quantitative approaches to solve existing or new issues.

2. Papers

The impact of the special issue comes from the submission of twenty-four research papers. From these twenty-four papers, only twelve were finally accepted for publication in this special collection. The contributions were selected based on their innovation and quality, demonstrating their applicability and importance in the field. A brief description of each selected paper for this collection is presented in the following.

In paper *"Normality index of ventricular contraction based on a statistical model from FADS"* by L. Jiménez-Ángeles et al., the authors present a probability density function model of the 3 most significant factors present in a dynamic series of equilibrium radionuclide angiography images extracted from factor analysis of dynamic structures for a control group which is presented. The authors also present an index, based on the likelihood between the control group's contraction model and a sample of normal subjects. The proposed index provides a measure, consistent with the phase analysis currently used in clinical environments.

Paper *"Development of an expert system as a diagnostic support of cervical cancer in atypical glandular cells, based on*

fuzzy logics and image interpretation" by K. R. Domínguez-Hernández, et al. describes the development of an expert system which can provides a diagnosis to cervical neoplasia precursors injuries through the integration of fuzzy logics and image interpretation techniques.

The paper, "*The iOSC3 system: using ontologies and SWRL rules for intelligent supervision and care of patients with acute cardiac disorders*" by M. Martínez-Romero et al., presents the design, development, and validation of iOSC3, an ontology-based system for intelligent supervision and treatment of critical patients with acute cardiac disorders.

In paper "*An intelligent system approach for asthma prediction in symptomatic preschool children*" by Chatzimichail et al., the authors present a study where a new method for asthma outcome prediction based on principal component analysis and least square support vector machine classifier is proposed.

In paper *"Ontology-oriented diagnostic system for traditional chinese medicine based on relation refinement"* by P. Gu et al., the authors define the diagnosis in traditional Chinese medicine as the discovery of fuzzy relations between symptoms and syndromes. The authors created an ontology-oriented diagnostic system to address the knowledge-based diagnosis through relation refinement based on a well-defined ontology of syndromes.

In paper *"Evaluation of the diagnostic power of thermography in breast cancer using Bayesian network classifiers,"* C.-R. Nicandro et al. evaluate the diagnostic power of thermography in breast cancer using Bayesian network classifiers. The authors show how the information provided by the thermal image can be used to characterize patients suspected of having cancer.

The paper *"Detection of structural changes in tachogram series for the diagnosis of atrial fibrillation events,"* by F. Ieva et al. presents a new statistical method to deal with the identification of atrial fibrillation events based on the order of identification of the ARIMA models used for describing the RR time series that characterize the different phases of atrial fibrillation (before-, during, and after-AF).

The paper *"Study of the effect of breast tissue density on detection of masses in mammograms,"* A. García-Manso et al. study the effect of BI-RADS density in their project for developing an image-based CAD system to detect masses in mammograms.

The paper *"Mobile personal health system for ambulatory blood pressure monitoring*" by L. J. Mena et al. presents ARV mobile, a multiplatform mobile personal health monitor (PHM) application for ambulatory blood pressure (ARP) monitoring that has the potential to aid in the acquisition and analysis of detailed profile of ABP and heart rate (HR), improving the early detection and intervention of hypertension, offering a tailored hypertension control, allowing also and detecting potential abnormal BP and HR levels for timely medical feedback.

The paper "*Statistical evaluation of a fully automated mammographic breast density algorithm*" M. by Abdolell et al. presents a study of the statistical evaluation of a fully automated, area-based mammographic density measurement algorithm. This algorithm is behind the idea that visual assessments of mammographic breast density by radiologist are used in clinical practice; however, these assessments have shown weaker associations with breast cancer risk than area-based quantitative methods.

The paper *"Locomotor development prediction based on statistical model parameters identification*" by A. Wildemann et al. presents and introduces an approach for parameter identification on a statistical prediction model by means of the use of available individual data. The application of the method was done for predicting the movement of a patient with congenital motility disorders.

Finally, the paper *"Design and customization of telemedicine systems"* by C. I. Martínez-Alcalá et al. is focused on the improvement of telemedicine systems with the aim of customizing therapies according to the profile and disability of patients. The work proposes the adoption of user-centered design methodology for the design and development of telemedicine systems in order to support the rehabilitation of patients with neurological disorders.

3. Conclusions

As can be seen, all the papers accepted have a direct relation with the scope of the special issue, and all of them provide quite interesting research techniques, models, and studies directly applied to the area of medical diagnostic systems.

Angel García-Crespo
Giner Alor-Hernandez
Linamara Battistella
Alejandro Rodríguez-González

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 264246, 10 pages
http://dx.doi.org/10.1155/2013/264246

Research Article

Evaluation of the Diagnostic Power of Thermography in Breast Cancer Using Bayesian Network Classifiers

Cruz-Ramírez Nicandro,[1] Mezura-Montes Efrén,[1] Ameca-Alducin María Yaneli,[1] Martín-Del-Campo-Mena Enrique,[2] Acosta-Mesa Héctor Gabriel,[1] Pérez-Castro Nancy,[3] Guerra-Hernández Alejandro,[1] Hoyos-Rivera Guillermo de Jesús,[1] and Barrientos-Martínez Rocío Erandi[1]

[1] *Departamento de Inteligencia Artificial, Universidad Veracruzana, Sebastián Camacho 5, Centro, 91000 Xalapa, VZ, Mexico*
[2] *Centro Estatal de Cancerología: Miguel Dorantes Mesa, Aguascalientes 100, Progreso Macuiltepetl, 91130 Xalapa, VZ, Mexico*
[3] *Laboratorio Nacional de Informática Avanzada (LANIA) A.C. Rébsamen 80, Centro, 91000 Xalapa, VZ, Mexico*

Correspondence should be addressed to Cruz-Ramírez Nicandro; ncruz@uv.mx

Received 25 October 2012; Revised 4 April 2013; Accepted 22 April 2013

Academic Editor: Alejandro Rodríguez González

Breast cancer is one of the leading causes of death among women worldwide. There are a number of techniques used for diagnosing this disease: mammography, ultrasound, and biopsy, among others. Each of these has well-known advantages and disadvantages. A relatively new method, based on the temperature a tumor may produce, has recently been explored: thermography. In this paper, we will evaluate the diagnostic power of thermography in breast cancer using Bayesian network classifiers. We will show how the information provided by the thermal image can be used in order to characterize patients suspected of having cancer. Our main contribution is the proposal of a score, based on the aforementioned information, that could help distinguish sick patients from healthy ones. Our main results suggest the potential of this technique in such a goal but also show its main limitations that have to be overcome to consider it as an effective diagnosis complementary tool.

1. Introduction

Breast cancer is one of the main causes of death among women worldwide [1]. Moreover, a specificity is required in the diagnosis of such a disease given that an incorrect classification of a sample as a false positive may lead to the surgical removal of the breast [2]. Nowadays, there are different techniques for carrying out the diagnosis: mammography, ultrasound, MRI, biopsies, and, more recently, thermography [3–6]. In fact, thermography started in 1956 [7] but was discarded some years later because of the poor quality of the thermal images [8] and the low specificity values it achieved. However, with the development of new thermal imaging technology, thermography has reappeared and is being seriously considered as a complementary tool for the diagnosis of breast cancer [9]. Because of specificity required, it is compulsory to have as many available tools as possible to reduce, on the one hand, the number of false positives and, on the other hand, to achieve high sensitivity. Although open biopsy is regarded as the gold standard technique for diagnosing breast cancer, it is practically the last diagnostic resource used since it is an invasive procedure that represents not only significant health implications but also psychological and economic ones also [10]. Other techniques, which are not necessarily invasive, have implicit risks or limitations such as X-ray exposure, interobserver interpretability and difficult access to high-tech expensive equipment [11, 12]. Thermography is also noninvasive, but it has the advantage of using a cheaper device (an infrared camera), which is far more portable than those used in mammography, MRI, and ultrasound. Furthermore, it can be argued that some of the variables considered by thermography may be more easily interpreted than those of some of the aforementioned techniques. As a matter of fact, in this paper we will explore and assess this argument in order to measure the potential of such a technique as a diagnostic tool for breast cancer. Moreover, our main contribution is

the proposal of a score, based not only on thermographic variables but also on variables that portray more information than temperature alone, that might help differentiate sick patients from healthy ones. We will also explore the potential of thermography in diagnosing women below the age of 50, which would allow the detection of the disease in its early stages, thus reducing the percentage of mortality.

The rest of the paper is divided as follows. In Section 2, we will present some related research that places our research in context and thus appreciates our contribution. In Section 3, we explain the materials and methods used in our experiments. In Section 4, we will present the methodology and the experimental results. In Section 5, we will discuss these results and, finally, in Section 6, we will conclude our paper and give directions regarding future research.

2. Related Research

In our review of the related literature, we divided these into three categories: introductory, image-based, and data-based works [13–17]. The introductory research mainly points out the potential of thermography as an alternative diagnostic tool for breast cancer comparing its performance to other diagnostic methods such as mammography and biopsy [18, 19]. Unfortunately, because this research is intended as an introduction to the topic, it lacks some important details about the data used in these studies as well as the analyses carried out.

The image-based works mainly range from cluster analyses applied to thermal images (to differentiate healthy from sick breasts) [20] to fractal analyses (to characterize the geometry of the malignant lesions) [21] to the camera calibration for capturing thermal images [3, 22].

The data-based investigations present statistical analyses of patient databases (healthy and sick) such as nonparametric tests, correlation, and analysis of variance; artificial intelligence analyses such as artificial neural networks and Bayesian analysis; and numerical models such as physical and simulation models (bioheat equations) [8, 9, 23–26]. Only a small number of papers propose a score formed from thermographic data [27, 28] but they only propose a maximum of 5 variables to form such a score. In our research, we propose 14 variables to calculate this score: this is the main contribution of the paper alongside the analysis of the diagnostic power of the proposed variables. In Section 3, we will present those variables in more detail and, in Section 4, we will evaluate how informative these variables are in the diagnosis of breast cancer. To end this section, it is important to mention that although the research in this category is very interesting, in some of them the methodology is not clear. This prevents one from easily reproducing the experiments carried out there. We have done our best to present a clear methodology so that our results can be reproduced.

3. Materials and Methods

3.1. The Database. For our experiments, we used a real-world database which was provided by an oncologist who has specialized in the study of thermography since 2008, consisting of 98 cases: 77 cases are patients with breast cancer (78.57%) and 21 cases are healthy patients (21.43%). All the results (either sick or healthy) were confirmed by an open biopsy, which is considered the gold standard diagnostic method for breast cancer [29]. We include in this study 14 explanatory variables (attributes): 8 of them form our score (proposed by the expert), 6 are obtained from the thermal image, one variable is the score itself, and the final variable is age which was discretized in three categories as this is recommended for the selected algorithms [30–32]. In Table 1, we give details of the name, definitions, and values of each of these variables. The dependent variable (class) is the outcome (cancer or no cancer).

3.2. Bayesian Networks. A Bayesian network (BN) [33, 34] is a graphical model that represents relationships of a probabilistic nature among variables of interest. Such networks consist of a qualitative part (structural model), which provides a visual representation of the interactions amid variables, and a quantitative part (set of local probability distributions), which permits probabilistic inference and numerically measures the impact of a variable or sets of variables on others. Both the qualitative and quantitative parts determine a unique joint probability distribution over the variables in a specific problem [33–35]. In other words, a Bayesian network is a directed acyclic graph consisting of [36]: (a) nodes (circles), which represent random variables; arcs (arrows), which represent probabilistic relationships among these variables and (b) for each node, there is a local probability distribution attached to it, which depends on the state of its parents.

Figures 3 and 4 (see Section 4) show examples of a BN. One of the great advantages of this model is that it allows the representation of a joint probability distribution in a compact and economical way by making extensive use of conditional independence, as shown in (1):

$$P\left(X_1, X_2, \ldots, X_n\right) = \prod_{i=1}^{n} P\left(X_i \mid Pa\left(X_i\right)\right), \tag{1}$$

where $Pa(X_i)$ represents the set of parent nodes of X_i, that is, nodes with arcs pointing to X_i. Equation (1) also shows how to recover a joint probability from a product of local conditional probability distributions.

3.2.1. Bayesian Network Classifiers. Classification refers to the task of assigning class labels to unlabeled instances. In such a task, given a set of unlabeled cases on the one hand and a set of labels on the other, the problem to solve lies in finding a function that suitably matches each unlabeled instance to its corresponding label (class). As can be inferred, the central research interest in this specific area is the design of automatic classifiers that can estimate this function from data (in our case, we are using Bayesian networks). This kind of learning is known as supervised learning [37–39]. For the sake of brevity and the lack of space, we have not written here the code of the 2 procedures used in the tests carried out in this research. We have only briefly described them and refer the reader to their original sources. The procedures used in these tests are

Table 1: Names, definitions, and values of variables. In the experiments the positive value is discretized to 1 and the negative value is discretized to 0. All the values of qualitative variables are given by the image analyst.

Variable name	Definition	Variable value	Variable type
Asymmetry	Temperature difference (in Celsius) between the right and the left breasts	If difference < 1°C, then value = 5, difference between 1°C and 2°C, the value is 10, and difference > 2°C, the value is 15	Nominal (5, 10, 15)
Thermovascular network	Number of veins with the highest temperature	If the visualization is abundant vascularity, the value is 15, if it is moderate, the value is 10, and if it is slight, the value is 5	Nominal (5, 10, 15)
Curve pattern	Heat area under the breast	If heat visualized is abundant, the value is 15, if it is moderate, the value is 10, and if it is slight, the value is 5	Nominal (5, 10, 15)
Hyperthermia	Hottest point of the breast	If there is at least one hottest point, the value is 20 and otherwise the value is 0	Binary (0, 20)
2c	Temperature difference between the hottest points of the two breasts	If difference between 1 and 10, the value is 10, difference between 11 and 15, the value is 15, difference between 16 and 20, the value is 20 and if difference > 20, the value is 25	Nominal (10, 15, 20, 25)
F unique	Amount of hottest points	If sum = 1, the value is 40, if sum = 2, the value is 20, if sum = 3, the value is 10, and if sum > 3, the value is 5	Nominal (5, 10, 20, 40)
1c	Hottest point in only one breast	If the hottest point is only one breast, the value is 40 and if the hottest point is both breasts, the value is 20	Binary (20, 40)
Furrow	Furrows under the breasts	If the furrow is visualized, the value is positive; if not,the value is negative	Binary (0, 1)
Pinpoint	Veins going to the hottest points of the breasts	If the veins are visualized, the value is positive; if not, the value is negative	Binary (0, 1)
Hot center	The center of the hottest area	If the center of the hottest is visualized, the value is positive; if not, the value is negative	Binary (0, 1)
Irregular form	Geometry of the hot center	If the hot center is visualized like a nongeometrical figure, the value is positive; if not, the value is negative	Binary (0, 1)
Histogram	Histogram in form of a isosceles triangle	If the histogram is visualized as a triangle form, the value is positive; if not, the value is negative	Binary (0, 1)
Armpit	Difference temperature between the 2 armpits	If the difference = 0, the value in both is negative; if not, the value is positive; consequently the other is negative	Binary (0, 1)
Breast profile	Visually altered profile	If an altered profile is visualized abundantly, the value is 3, if it is moderate, value is 2, if it is small, the value is 1, and if it does not exist, the value is 0	Binary (0, 1)
Score	The sum of values of the previous 14 variables	If the sum < 160, then the value is negative for cancer; if the sum ≥ 160, the value is positive for cancer	Binary (0, 1)
Age	Age of patient	If the age < 51, the value is 1, if the age between 51 and 71, the value is 2, and if age > 71, the value is 3	Binary (0, 1)
Outcome	The result is obtained via open biopsy	The values are cancer or no-cancer	Binary (0, 1)

(a) the Naïve Bayes classifier, (b) Hill-Climber, and (c) Repeated Hill-Climber [38, 40, 41].

(a) The Naïve Bayes classifier (NB) is one of the most effective classifiers [38] and the benchmark against which state-of-the-art classifiers have to be compared. Its main appeals lie in its simplicity and accuracy: although its structure is always fixed (the class variable has an arc pointing to every attribute), it has been shown that this classifier has a high classification accuracy and optimal Bayes's error (see Figure 3, Section 4). In simple terms, the NB learns, from a

training data sample, the conditional probability of each attribute given the class. Then, once a new case arrives, the NB uses Bayes's rule to compute the conditional probability of the class given the set of attributes selecting the value of the class with the highest posterior probability.

(b) Hill-Climber is a Weka's [41] implementation of a search and scoring algorithm, which uses greedy-hill-climbing [42] for the search part and different metrics for the scoring part, such as Bayesian information criterion (BIC), Bayesian Dirichlet (BD), Akaike information criterion (AIC), and minimum description length (MDL) [43]. For the experiments reported here, we selected the MDL metric. This procedure takes an empty graph and a database as input and applies different operators for building a Bayesian network: addition, deletion, or reversal of an arc. In every search step, it looks for a structure that minimizes the MDL score. In every step, the MDL is calculated and procedure Hill-Climber keeps the structure with the best (minimum) score. It finishes searching when no new structure improves the MDL score of the previous network.

(c) Repeated Hill-Climber is a Weka's [41] implementation of a search and scoring algorithm, which uses repeated runs of greedy hill-climbing [42] for the search part and different metrics for the scoring part, such as BIC, BD, AIC, and MDL. For the experiments reported here, we selected the MDL metric. In contrast to the simple Hill-Climber algorithm, Repeated Hill-Climber takes as input a randomly generated graph. It also takes a database and applies different operators (addition, deletion, or reversal of an arc) and returns the best structure of the repeated runs of the Hill-Climber procedure. With this repetition of runs, it is possible to reduce the problem of getting stuck in a local minimum [35].

3.3. Evaluation Method: Stratified k-Fold Crossvalidation. We followed the definition of the crossvalidation method given by Kohavi [37]. In k-fold crossvalidation, we split the database D in k mutually exclusive random samples called the folds: $D_1, D_2, \ldots, D_k$, where said folds have approximately the same size. We trained this classifier each time $i \in 1, 2, \ldots, k$ using $D \setminus D_i$ and testing it on D_i (again, the symbol denotes set difference). The crossvalidation accuracy estimation is the total number of correct classifications divided by the sample size (total number of instances in D). Thus, the k-fold crossvalidation estimate is as follows:

$$\mathrm{acc}_{\mathrm{cv}} = \frac{1}{n} \sum_{(v_i, y_i) \in D} \delta\left(I\left(D \setminus D_{(i)}, v_i\right), y_i\right), \tag{2}$$

where $(I(D \setminus D_{(i)}, v_i), y_i)$ denotes the label assigned by inducer I to an unlabeled instance v_i on dataset $D \setminus D_{(i)}$, y_i is the class of instance v_i, n is the size of the complete dataset, and $\delta(i, j)$ is a function where $\delta(i, j) = 1$ if $i = j$ and 0 if $i \neq j$. In other words, if the label assigned by the inducer to the unlabeled instance v_i coincides with class y_i, then the result is 1; otherwise, the result is 0; that is, we consider a 0/1 loss function in our calculations of (2). It is important to mention that in stratified k-fold crossvalidation, the folds contain approximately the same proportion of classes as in the complete dataset D. A special case of crossvalidation occurs when $k = n$ (where n represents the sample size). This case is known as leave-one-out crossvalidation [37, 39].

For both evaluation methods, we assessed the performance of the classifiers presented in Section 3.2 using the following measures [44–47].

(a) Accuracy: the overall number of correct classifications divided by the size of the corresponding test set:

$$a = \frac{\mathrm{cc}}{n}, \tag{3}$$

where cc represents the number of cases correctly classified and n is the total number of cases in the test set.

(b) Sensitivity: the ability to correctly identify those patients who actually have the disease:

$$S = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{4}$$

where TP represents true positive cases and FN is false negative cases.

(c) Specificity: the ability to correctly identify those patients who do not have the disease:

$$Sp = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{5}$$

where TN represents true negative cases and FP is false positive cases.

4. Methodology and Experimental Results

We used stratified 10-fold crossvalidation on the 98-case database described in Section 3.1. All the algorithms described in Section 3.2.1 used this data in order to learn a classification model. Once we have this model, we then evaluate its performance in terms of accuracy, sensitivity, and specificity. We used Weka [41] for the tests carried out here (see their parameter set in Table 2). For comparison purposes other classifiers were included: a multilayer perceptron (MLP) neural network and decision trees (ID3 and C4.5) with default parameters. The fundamental goal of this experiment was to assess the diagnostic power of the thermographic variables that form the score and the interactions among these variables. To illustrate how the variable values are obtained, we cite one example.

(a) In Figure 1 we show the type of images obtained by the thermal imager; in this case, the front of the breast thermography. Using *ThermaCAM Researcher Professional 2.9* [48] software, we detect the hottest areas of the breast that pass from red to gray. The breast whose furrow displays the largest gray area is assigned a positive value and the other a negative one.

Table 2: Parameter values for Hill-Climber and Repeated Hill-Climber.

Parameters	Hill-Climber	Repeated Hill-Climber
The initial structure NB (Naïve Bayes)	False	False
Number of parents	100,000	100,000
Runs	—	10
Score type	MDL	MDL
Seed	—	1
Arc reversal	True	True

Table 3: Accuracy, sensitivity, and specificity results for the three Bayesian network classifiers presented in Section 3.2.1.

	Naïve Bayes	Hill-Climber	Repeated Hill-Climber
Accuracy	71.88% (±12.61)	76.10% (±7.10)	76.12% (±7.19)
Sensitivity	82% (74–91)	97% (94–100)	99% (96–100)
Specificity	37% (15–59)	0% (0-0)	0% (0-0)

Table 4: Accuracy, sensitivity, and specificity of artificial neural network, decision trees ID3 and C4.5 for the thermography.

	Artificial neural network	Decision tree ID3	Decision tree C4.5
Accuracy	67.47% (±15.65)	73.19% (±12.84)	75.50% (±6.99)
Sensitivity	82% (73–91)	87% (79–94)	94% (88–99)
Specificity	33% (13–53)	29% (9–48)	0% (0-0)

Table 5: Confusion matrix of Naïve Bayes.

	Cancer	Noncancer	Total
Cancer	TP 65	FN 12	77
Noncancer	FP 14	TN 7	21
			98

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

Table 6: Confusion matrix of Hill-Climber.

	Cancer	Non-cancer	Total
Cancer	TP 75	FN 2	77
Non-cancer	FP 21	TN 0	21
			98

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

In Figure 2 we show a general overview of the procedure of breast thermography, from thermal image acquisition to the formation of the score.

Tables 3, 4, 5, 6, 7, 8, 9, and 10 show the numerical results of this experiment. Figures 3 and 4 show the structures resulting from running Hill-Climber and Repeated Hill-Climber classifiers and Figure 5 shows the decision tree (C4.5). We do not present the structure of the Naïve Bayes classifier since it is always fixed: there is an arc pointing to every attribute from the class. For the accuracy test, the standard deviation is shown next to the accuracy result. For the remaining tests, their respective 95% confidence intervals (CI) are shown in parentheses.

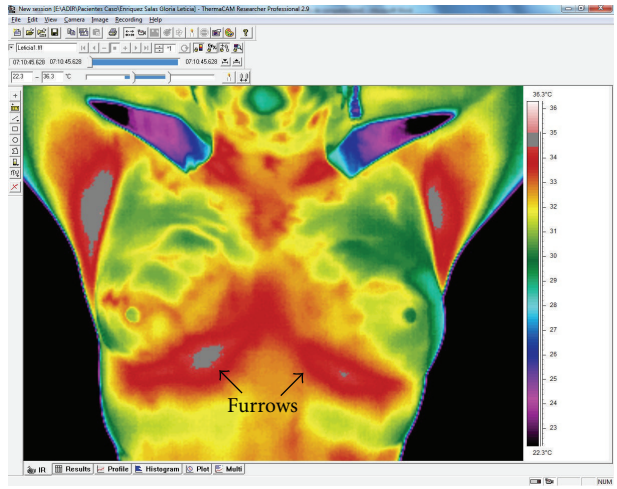


Figure 1: Thermal image showing the temperature of the color-coded breasts. The red and gray tones represent hotter areas.

Table 7: Confusion matrix of Repeated Hill-Climber.

	Cancer	Non-cancer	Total
Cancer	TP 76	FN 1	77
Non-cancer	FP 21	TN 0	21
			98

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

Table 8: Confusion matrix of artificial neural network.

	Cancer	Non-cancer	Total
Cancer	TP 58	FN 19	77
Non-cancer	FP 15	TN 6	21
			98

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

Table 9: Confusion matrix of decision tree ID3.

	Cancer	Non-cancer	Total
Cancer	TP 67	FN 10	77
Non-cancer	FP 15	TN 6	21
			98

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

Table 10: Confusion matrix of decision tree C4.5.

	Cancer	Non-cancer	Total
Cancer	TP 76	FN 1	77
Non-cancer	FP 21	TN 0	21
			98

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

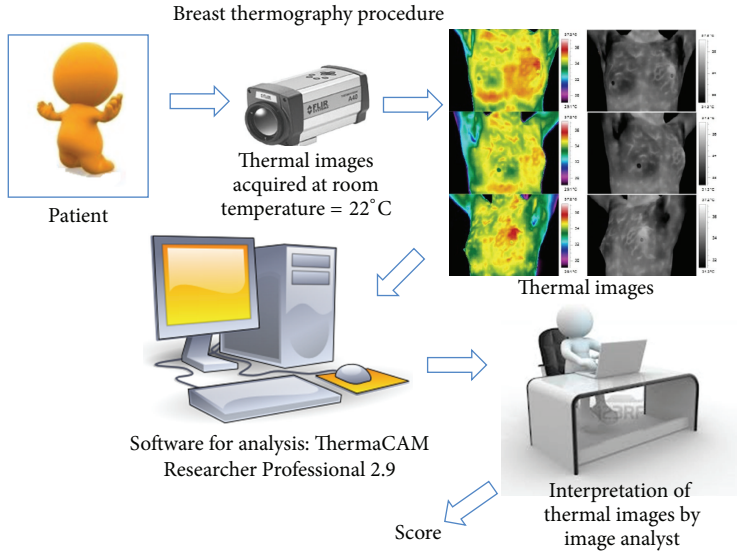


Figure 2: Breast thermography procedure.

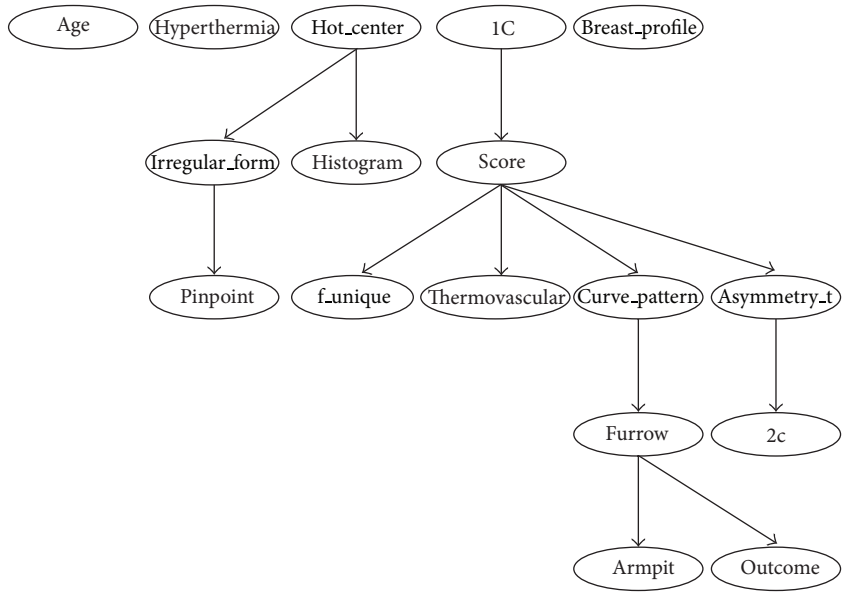


Figure 3: Bayesian network built by procedure of Hill-Climber using the 98-case database. Only variable furrow is directly related to the outcome. Once the variable furrow is known, all the other variables are independent of the class.

5. Discussion

The main objective of this paper is to assess the diagnostic power of thermography in breast cancer using Bayesian network classifiers. As can be seen from Table 3, the overall accuracy is still far from a desirable value. We chose Bayesian networks for the analyses because this model does not only carry out a classification task but it is also able to show interactions between the attributes and the class as well as interactions among the attributes themselves. This ability of Bayesian networks allows us to visually identify which attributes have a direct influence over the outcome and how they are related to one another. The MLP shows a comparable performance but lacks the power of explanation: it is not possible to query this network to know how it reached a specific decision. On the other hand, decision trees do have this explanation capability but lack the power to represent interactions among attributes (explanatory variables). Figures 3 and 4 depict that only 5 variables (out of 16) are directly related to the score: 1C, f_unique, thermovascular, curve_pattern, and asymmetry_t. Hence we can see that the score influence on the class *outcome* is null and the variable *furrow* (this variable is part of the score) is the only one that affects the class. Figure 5 shows that procedure C4.5 also identifies 2 of those 5 variables as being

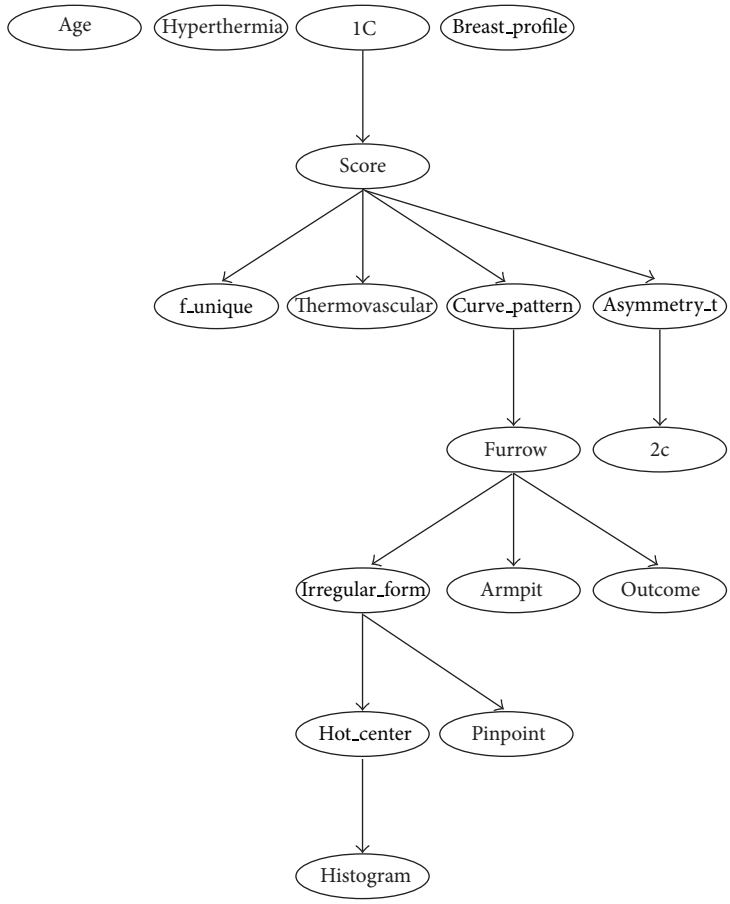


Figure 4: Bayesian network built by procedure of Repeated Hill-Climber using the 98-case database. Only variable furrow is directly related to the outcome. Once the variable furrow is known, all the other variables are independent of the class.

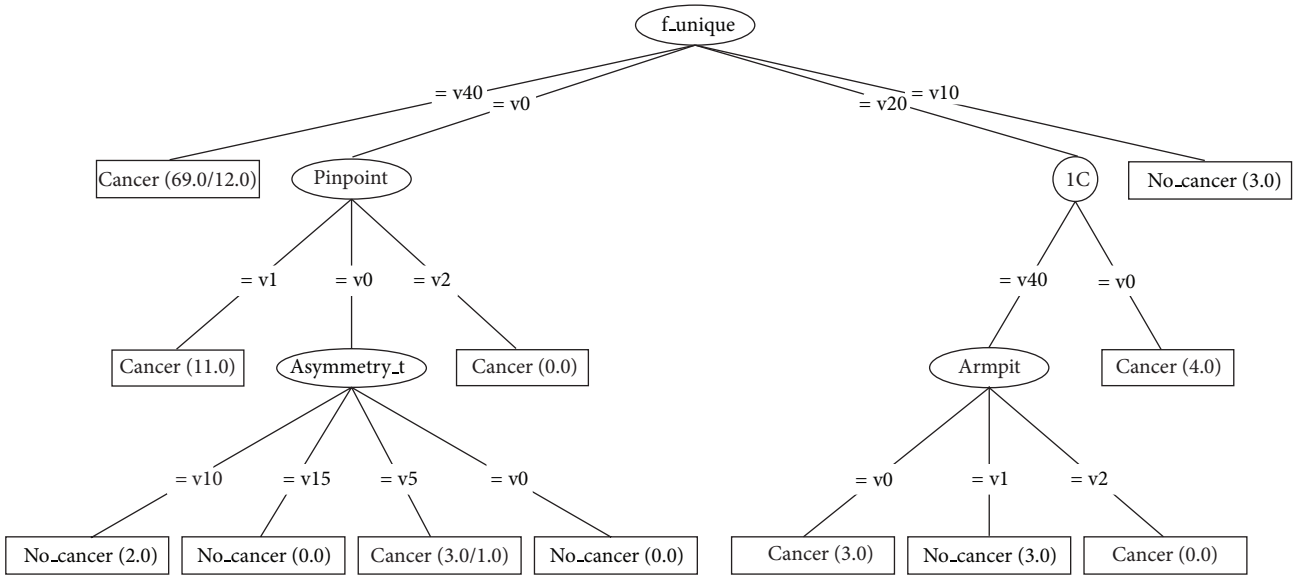


Figure 5: Decision tree C4.5 using the 98-case database.

the most informative ones for making a decision: f_unique and asymmetry_t. In fact, if we only consider these attributes, we get the same classification performance as that when taking into account all thermographic variables. Other models, such as artificial neural networks, cannot easily identify this situation. As seen in Section 3.2, the extensive use of conditional independence allows Bayesian networks to potentially disregard spurious causes and to easily identify direct influences from indirect ones. In other words, once these variables are known, they render the rest of the variables independent from the outcome. Another surprising result is that of variable age: some other tests consider this to be an important observation for the diagnosis of breast cancer [30–32]. However, our analyses suggest that, at least with the database used in our experiments, age is not important in a diagnosis when using thermography. As can be seen from Figures 3 and 4, age is disconnected from the rest of the variables. This may imply that thermography shows potential for diagnosing breast cancer in women younger than 50 years of age.

Regarding the sensitivity performance of our models (see Table 3), Hill-Climber and Repeated Hill-Climber achieve a perfect value of 100%. This means that, at least with our database, thermography is excellent for identifying sick patients. Naïve Bayes classifier shows a significantly worse performance; it can be argued that this performance is due to the noise that the rest of the variables may add. Once again, if we only considered the 5 variables mentioned above, we would get the same results as those using Hill-Climber and Repeated Hill-Climber. Other models would not be capable of revealing this situation. Of course, it is mandatory to get more data in order to confirm such results.

It is important to point out that the Hill-Climber and Repeated Hill-Climber procedures identify the same 5 variables as directly influencing the outcome.

Regarding the specificity performance of our models (see Table 3), Hill-Climber and Repeated Hill-Climber achieve the worst possible value of 0%. This means these 5 variables, while being informative when detecting the presence of the disease, are not useful for detecting the absence of such disease (see Tables 5–10). On the other hand, the noise that the rest of the attributes produce when detecting the disease seems to work the other way around: it is not noise but information that makes Naïve Bayes achieve a specificity of 33%. Of course, such a value is far from desirable, but this result makes us think of proposing two different scores (one for sensitivity and one for specificity) with two different sets of variables. But our proposal of a score is a first approximation to combine thermographic variables in such a way as to allow us to tell sick patients from healthy ones. Our results show that such a score needs to be refined in order to more easily identify these types of patients.

Although the results may be discouraging, we strongly believe that they are a step forward in order to more deeply comprehend the phenomenon under investigation: breast cancer. In fact, we have proposed a score that takes into account more information than just that of temperature. Until now, few areas of research have considered other variables apart from that of temperature [27, 28]. Those papers include in their analyses a total of 5 variables that can be extracted from the information a thermogram provides. Our score includes 16 variables and our work, to the best of our knowledge, presents the first analysis of this kind of data using Bayesian networks. What this analysis suggests is a refinement of the score, probably in the sense of proposing a more complex function to represent it beyond the simple addition of the values of each attribute. Intuitively, we thought that other variables, such as hyperthermia or thermovascular network, would be more significant in differentiating sick patients from healthy ones.

In the case of the database, we are aware of the limitations regarding the number of cases and the imbalance of the number of classes. Thus, we would need to collect more data so that more exhaustive tests can be carried out.

6. Conclusions and Future Work

Thermography has been used as an alternative method for the diagnosis of breast cancer since 2005. The basic principle is that lesions in the breasts are hotter than healthy regions. In our experience, only taking into account temperature is not enough to diagnose breast cancer. That is why we proposed a score that considers more information than only temperature alone. We have found that only 5 attributes that are part of this score are the unique direct influence needed to determine if a patient has cancer.

Although some other research projects show better performance than ours, their methodology to carry out the experiments is not clear; thus these experiments cannot be reproduced. Therefore, we need to more closely explore the details of these models and the nature of their data. In this paper we have done our best to present the methodology used in our experiments as clear as possible so that they indeed can be reproduced. It is true that we do not give details about how the database was formed (since this is not the primary goal of the paper). However, we believe that if we make this database available, researchers who want to reproduce our experiments should be able to do so without much trouble.

We have found that the framework of Bayesian networks provides a good model for analyzing this kind of data: it can visually show the interactions between attributes and outcome as well as the interactions among attributes and numerically measure the impact of each attribute on the class.

Although we obtained excellent sensitivity results, we also obtained very poor specificity results. The sensitivity values are consistent with the expectations of the expert, and a discussion about the helpfulness of the Bayesian network is already underway in order to better understand the disease. Given that breast cancer has a special requirement of specificity values, we have to more deeply investigate the causes of those poor results. One possible direction for future research is to collect more balanced data using techniques such as SMOTE [49], ADASYN [50], AdaC1 [51], and GSVM-RU [52]. Another possible direction is to design a more complex score that includes a more complex function compared to that of a simple sum. A third direction we can detect is reviewing how the variables are collected and try to reduce subjectivity in them. Finally, we have also detected that medical doctors usually take into account more information than that

supplied to the models for diagnosing breast cancer. Thus, we can also work more in the area of knowledge elicitation.

Conflict of Interests

The authors declare that they have no conflict of interests.

References

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *A Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[2] J. Gnerlich, D. B. Jeffe, A. D. Deshpande, C. Beers, C. Zander, and J. A. Margenthaler, "Surgical removal of the primary tumor increases overall survival in patients with metastatic breast cancer: analysis of the 1988–2003 SEER data," *Annals of Surgical Oncology*, vol. 14, no. 8, pp. 2187–2194, 2007.

[3] E. Y. K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor," *International Journal of Thermal Sciences*, vol. 48, no. 5, pp. 849–859, 2009.

[4] J. Bonnema, A. N. Van Geel, B. Van Ooijen et al., "Ultrasound-guided aspiration biopsy for detection of nonpalpable axillary node metastases in breast cancer patients: new diagnostic method," *World Journal of Surgery*, vol. 21, no. 3, pp. 270–274, 1997.

[5] M. D. Schnall, J. Blume, D. A. Bluemke et al., "MRI detection of distinct incidental cancer in women with primary breast cancer studied in IBMC 6883," *Journal of Surgical Oncology*, vol. 92, no. 1, pp. 32–38, 2005.

[6] B. M. Geller, K. Kerlikowske, P. A. Carney et al., "Mammography surveillance following breast cancer," *Breast Cancer Research and Treatment*, vol. 81, no. 2, pp. 107–115, 2003.

[7] K. R. Foster, "Thermographic detection of breast cancer," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 6, p. 10, 1998.

[8] G. C. Wishart, M. Campisi, M. Boswell et al., "The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy," *European Journal of Surgical Oncology*, vol. 36, no. 6, pp. 535–540, 2010.

[9] N. Arora, D. Martins, D. Ruggerio et al., "Effectiveness of a non-invasive digital infrared thermal imaging system in the detection of breast cancer," *American Journal of Surgery*, vol. 196, no. 4, pp. 523–526, 2008.

[10] H. M. Verkooijen, P. H. M. Peeters, E. Buskens et al., "Diagnostic accuracy of large core needle biopsy for nonpalpable breast disease: a meta-analysis," *British Journal of Cancer*, vol. 82, no. 5, pp. 1017–1021, 2000.

[11] C. K. Kuhl, S. Schrading, C. C. Leutner et al., "Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer," *Journal of Clinical Oncology*, vol. 23, no. 33, pp. 8469–8476, 2005.

[12] M. Kriege, C. T. M. Brekelmans, C. Boetes et al., "Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition," *New England Journal of Medicine*, vol. 351, no. 5, pp. 427–519, 2004.

[13] F. Gutierrez, J. Vazquez, L. Venegas et al., "Feasibility of thermal infrared imaging screening for breast cancer in rural communities of southern mexico: the experience of the centro de estudios y prevencion del cancer (ceprec)," in *Proceedings of the 2009 ASCO Annual Meeting*, p. 1521, American Society of Clinical Oncology, 2009.

[14] Q. Hairong, T. K. Phani, and L. Zhongqi, "Early detection of breast cancer using thermal texture maps," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 309–312, 2002.

[15] S. Ohsumi, S. Takashima, K. Aogi, and H. Usuki, "Prognostic value of thermographical findings in patients with primary breast cancer," *Breast Cancer Research and Treatment*, vol. 74, no. 3, pp. 213–220, 2002.

[16] E. Y. K. Ng, Y. Chen, and L. N. Ung, "Computerized breast thermography: study of image segmentation and temperature cyclic variations," *Journal of Medical Engineering and Technology*, vol. 25, no. 1, pp. 12–16, 2001.

[17] E. Y. K. Ng and N. M. Sudharsan, "Parametric optimization for tumour identification: bioheat equation using ANOVA and the Taguchi method," *Journal of Engineering in Medicine*, vol. 214, no. 5, pp. 505–512, 2000.

[18] E. E. Sterns and B. Zee, "Thermography as a predictor of prognosis in cancer of the breast," *Cancer*, vol. 67, no. 6, pp. 1678–1680, 1991.

[19] E. Y. K. Ng, L. N. Ung, F. C. Ng, and L. S. J. Sim, "Statistical analysis of healthy and malignant breast thermography," *Journal of Medical Engineering and Technology*, vol. 25, no. 6, pp. 253–263, 2001.

[20] M. EtehadTavakol, S. Sadri, and E. Y. K. Ng, "Application of K- and fuzzy c-means for color segmentation of thermal infrared breast images," *Journal of Medical Systems*, vol. 34, no. 1, pp. 35–42, 2010.

[21] M. EtehadTavakol, C. Lucas, S. Sadri, and E. Y. K. Ng, "Analysis of breast thermography using fractal dimension to establish possible difference between malignant and benign patterns," *Journal of Healthcare Engineering*, vol. 1, no. 1, pp. 27–44, 2010.

[22] Z. Damnjanović, D. Petrović, R. Pantović, and Z. Smiljanić, "Infra red digital imaging in medicine," *International Journal of Collaborative Research on Internal Medicine and Public Health*, vol. 2, no. 12, pp. 425–434, 2010.

[23] E. Y. K. Ng and S. C. Fok, "A framework for early discovery of breast tumor using thermography with artificial neural network," *Breast Journal*, vol. 9, no. 4, pp. 341–343, 2003.

[24] E. Y. K. Ng and N. M. Sudharsan, "Effect of blood flow, tumour and cold stress in a female breast: a novel time-accurate computer simulation," *Journal of Engineering in Medicine*, vol. 215, no. 4, pp. 393–404, 2001.

[25] E. Y. K. Ng and N. M. Sudharsan, "Numerical computation as a tool to aid thermographic interpretation," *Journal of Medical Engineering and Technology*, vol. 25, no. 2, pp. 53–60, 2001.

[26] E. Y. K. Ng, S. C. Fok, Y. C. Peh, F. C. Ng, and L. S. J. Sim, "Computerized detection of breast cancer with artificial intelligence and thermograms," *Journal of Medical Engineering and Technology*, vol. 26, no. 4, pp. 152–157, 2002.

[27] J. Wang, K. J. Chang, C. Y. Chen et al., "Evaluation of the diagnostic performance of infrared imaging of the breast: a preliminary study," *BioMedical Engineering*, vol. 9, p. 3, 2010.

[28] J. Wang, T. T. F. Shih, R. F. Yen et al., "The Association of Infrared Imaging Findings of the Breast with Hormone Receptor and Human Epidermal Growth Factor Receptor 2 Status of Breast Cancer," *Academic Radiology*, vol. 18, no. 2, pp. 212–219, 2011.

[29] F. Sardanelli, G. M. Giuseppetti, P. Panizza et al., "Sensitivity of MRI versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard," *American Journal of Roentgenology*, vol. 183, no. 4, pp. 1149–1157, 2004.

[30] S. S. Cross, T. J. Stephenson, T. Mohammed, and R. F. Harrison, "Validation of a decision support system for the cytodiagnosis of fine needle aspirates of the breast using a prospectively collected dataset from multiple observers in a working clinical environment," *Cytopathology*, vol. 11, no. 6, pp. 503–512, 2000.

[31] S. S. Cross, A. K. Dubé, J. S. Johnson et al., "Evaluation of a statistically derived decision tree for the cytodiagnosis of fine needle aspirates of the breast (FNAB)," *Cytopathology*, vol. 9, no. 3, pp. 178–187, 1998.

[32] A. J. Walker, S. S. Cross, and R. F. Harrison, "Visualisation of biomedical datasets by use of growing cell structure networks: a novel diagnostic classification technique," *The Lancet*, vol. 354, no. 9189, pp. 1518–1521, 1999.

[33] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann series in representation and reasoning, Morgan Kaufmann Publishers, Burlington, Mass, USA, 1988.

[34] L. G. Neuberg, "Causality: models, reasoning, and inference, by judea pearl, cambridge university press, 2000," *Econometric Theory*, vol. 19, pp. 675–685, 2003.

[35] N. Friedman and M. Goldszmidt, *Learning Bayesian Networks from Data*, University of California, Berkeley, Calif, USA; Stanford Research Institute, Menlo Park, Calif, USA, 1998.

[36] G. Cooper, "An overview of the representation and discovery of causal relationships using bayesian networks," in *Computation Causation Discovery*, 1999.

[37] R. Kohavi, *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*, Morgan Kauffmann, Boston, Mass, USA, 1995.

[38] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.

[39] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series in Data Management Systems, Elsevier, New York, NY, USA, 2006.

[40] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2001.

[41] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Series in Data Management Systems, Morgan Kauffmann, Boston, Mass, USA, 2nd edition, 2005.

[42] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 2009.

[43] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing Bayesian network structure learning algorithm," *Machine Learning*, vol. 65, no. 1, pp. 31–78, 2006.

[44] N. Lavrač, "Selected techniques for data mining in medicine," *Artificial Intelligence in Medicine*, vol. 16, no. 1, pp. 3–23, 1999.

[45] S. S. Cross, A. K. Dubé, J. S. Johnson et al., "Evaluation of a statistically derived decision tree for the cytodiagnosis of fine needle aspirates of the breast (FNAB)," *Cytopathology*, vol. 9, no. 3, pp. 178–187, 1998.

[46] S. S. Cross, T. J. Stephenson, T. Mohammed, and R. F. Harrison, "Validation of a decision support system for the cytodiagnosis of fine needle aspirates of the breast using a prospectively collected dataset from multiple observers in a working clinical environment," *Cytopathology*, vol. 11, no. 6, pp. 503–512, 2000.

[47] S. S. Cross, J. Downs, P. Drezet, Z. Ma, and R. F. Harrison, *Which Decision Support Technologies are Appropriate for the Cytodiagnosis of Breast Cancer?* World Scientific, Hackensack, NJ, USA, 2000.

[48] Inc FLIR System, "Thermacam researcher professional 2.9," 2009, http://support.flir.com/DocDownload/Assets/47/English/T5590091387.

[49] K. W. Bowyer, N. V. Chawla, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[50] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: adaptive synthetic sampling approach for imbalanced learning," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '08)*, pp. 1322–1328, June 2008.

[51] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, 2007.

[52] Y. Tang and Y. Q. Zhang, "Granular SVM with repetitive undersampling for highly imbalanced protein homology prediction," in *Proceedings of the IEEE International Conference on Granular Computing*, pp. 457–460, May 2006.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 618025, 16 pages
http://dx.doi.org/10.1155/2013/618025

Research Article

Design and Customization of Telemedicine Systems

Claudia I. Martínez-Alcalá,[1] Mirna Muñoz,[2] and Josep Monguet-Fierro[1]

[1] *Multimedia Engineering Department, Universitat Politècnica de Catalunya, ETSEIB, Avenida Diagonal 647, Barcelona, Spain*
[2] *Centro de Investigación en Matemáticas, Avenida Universidad No. 222, 98068 Zacatecas, ZAC, Mexico*

Correspondence should be addressed to Claudia I. Martínez-Alcalá; isabel.martinez.alcala@estudiant.upc.edu

Received 22 January 2013; Accepted 22 March 2013

Academic Editor: Giner Alor-Hernández

In recent years, the advances in information and communication technology (ICT) have resulted in the development of systems and applications aimed at supporting rehabilitation therapy that contributes to enrich patients' life quality. This work is focused on the improvement of the telemedicine systems with the purpose of customizing therapies according to the profile and disability of patients. For doing this, as salient contribution, this work proposes the adoption of user-centered design (UCD) methodology for the design and development of telemedicine systems in order to support the rehabilitation of patients with neurological disorders. Finally, some applications of the UCD methodology in the telemedicine field are presented as a proof of concept.

1. Introduction

In recent years, the advances in information and communication technology (ICT) have enabled the development of systems and applications aimed at supporting rehabilitation therapy and, therefore, contributing to the enrichment of patients' life quality [1–3]. The creation and implementation of web-enabled communication, patient services, and other e-Health initiatives have been significantly developed and enhanced in order to improve the quality of health services and maintain a competitive advantage. Consequently, the quality of health care has significantly improved [4, 5]. Traditionally, technology has supported health professionals by providing instruments, diagnosis, and different therapeutic treatments [6]. Moreover, the information and communication technologies have expanded their application to management and planning activities of health areas. Thus, companies involved in the health area must expand their capabilities to all stakeholders, including patients and the public in general towards robust, efficient, and friendly telemedicine systems [7, 8].

According to the literature [9, 10], doctors and nurses make use of the Internet in two mainly ways: (1) for communication, to send information through email, and (2) as an extensive library, to consult the clinical information. Also, it is mentioned that they have good computer skills, a positive attitude towards using the computer and Internet, and are motivated to use both ways on daily activities. However, [11] mentioned that some health professionals still show some resistance towards the acceptance of new technologies, even when some health sectors are beginning to integrate ICT in some of their fields.

Besides, Bernard et al. [12] mention that ICTs offer practical and timely mechanisms for continuing medical education allowing the improvement of educational programs for health professionals in rural areas [13–15]. Simultaneously, ICT may also have an important role in transferring clinical data [16]. The American Psychiatric Association (APA) states that once the information is stored, it is essential to have access to it. Moreover, recent technological advances have enabled the introduction of a broad range of telemedicine applications, such as teleradiology, teleconsultation, telesurgery, remote patient monitoring, and health-care records management that are supported by computer networks and wireless communication [17–20].

Through the development of user interfaces for health-care applications, researchers have empirically evaluated the effectiveness of diverse user-centered design (UCD) approaches [21, 22]. Health-care software developers often overlook relevant user features, user tasks, user preferences, and

usability issues. Thus, systems can provoke a decrease in productivity or simply be unusable. Several factors could be attributed to developing poor systems, such as cost and time restrictions and the lack of developers with sufficient knowledge on user-centered design [23].

According to the literature [24, 25], only 61% of information system projects meet the customer requirement specifications. Furthermore, 63% of projects exceed their estimated budgets due to the inadequate initial user analysis [24]. Incorporating good design principles in the beginning phase of a project not only saves time and cost, but also decreases changes in design late in the development process [25].

According to Wallach and Scholz [26], a UCD is defined as a design process and evaluation that pays attention to the intended user, focusing on what they will do with the product, where they will use it, and what features they consider essential. In particular, a UCD may be used by designers to address the needs of the patients and specialists about questions related to experiences with mental models of illness and disabilities. Figure 1 provides a graphical illustration of the process of UCD used in this research work.

As Figure 1 shows, this research presents a user centered design framework (UCD), specifically for customizing the design of health interfaces. The UCD contains four steps to follow: (1) analysis, (2) design, (3) implementation, and (4) evaluation. The goal of customizing the design of the interfaces is to model and develop systems based on the user needs and features. Therefore, the UCD used proposes to involve patients and specialists throughout the design phases, leading to easy-to-learn systems that increase user productivity and satisfaction and user acceptance and that reduce user errors and user training time.

2. State of the Art

Broadly, telemedicine refers to the use of information and telecommunication technologies to distribute information and/or expertise necessary for healthcare service provision, collaboration, and/or delivery among geographically separated participants, including physicians and patients [27, 28]. Different definitions highlight that telemedicine is an open and constantly evolving science, as it incorporates new technological advancements, and it responds and adapts to the changing societies' health needs [29]. Telemedicine supports different types of relationship between two or more actors who are not in a common physical space. The most common relationships established are (1) professional-professional, (2) professional-professional-patient, and (3) professional-patient.

Telemedicine covers different forms of information: (1) transmission (voice, sound, video, still picture, and text); (2) communication technologies (standard telephone lines, coaxial cable, satellite, microwave, digital wireless, ISDN, and Internet); and (3) user interfaces (desktop computers, laptop computers, personal digital assistants, fax machines, telephones, mobile phones, videophones, various standalone systems, and peripheries). All of them allow to carry out a wide range of activities, such as store-and-forward applications, which involve the asynchronous transmission of medical information, patient/health provider communications, and other data, and live audiographic encounters, which combine sound with still pictures, and perhaps, most importantly, the live two-way interactive video consultation [30, 31].

One of the main motivations for the application of ICT in both healthcare organizations, public and private, lies in the necessity of improving the information and providing medical care to a multitude of geographically dispersed agents. Clinical studies have shown that telemedicine is safe and cost-effective, compared with hospital treatment, especially with patients suffering from chronic diseases [32]. Besides, it is important to highlight that the introduction of telemedicine services must overcome a series of obstacles such as acceptance by patients, accessibility issues, technology costs, physical and psychological disabilities of patients, and acceptance and availability of medical personal [33, 34]. Despite all obstacles and failures, telemedicine concepts have been considered of great potential to support healthcare in particular for patients with neurological diseases. Since most of these patients are from older age groups, it is important to develop concepts, systems, and devices that can be handled by older patients and customized to their individual needs and limitations.

In this context, it is important to mention that despite its growth, there is a general feeling that telemedicine has a long way to go before it reaches its maximum potential [35]. Therefore, many different telemedical systems have been developed and used on various scales simply as experimental until a broader routine is carried out. Each of these scale systems has been more or less successful. However, the development of telemedicine systems and devices has often been defined by technical possibilities rather than by the needs of patients or their caregivers. Then, in order to know the performance of a telemedicine system, it needs to be evaluated not only for their benefit, but also for their feasibility, acceptance, and economic efficiency. Moreover, studies focused on telemedicine should consider research scenarios close to rehabilitation and should include older patients and patients with cognitive and physical limitations [35].

2.1. Advantages and Disadvantages with respect to Other Systems Developed. Table 1 summarizes the major advantages and disadvantages of other existing telemedicine systems.

3. Materials and Methods

3.1. Reference Model for the User-Centered Design of the Telemedicine Systems. The design process employed in the development of the telemedicine systems was inspired on the user-centered design (UCD) approach, a widely accepted methodology for creating usable applications or systems, which aims to truly meet the needs of users [3]. Figure 1 gives an overview of the developed method. The method is composed of four phases: (1) analysis, (2) design, (3) implementation, and (4) evaluation. The method phases are further described below.

The main challenge of this method is the customization of activities according to the user's needs. Besides, this approach

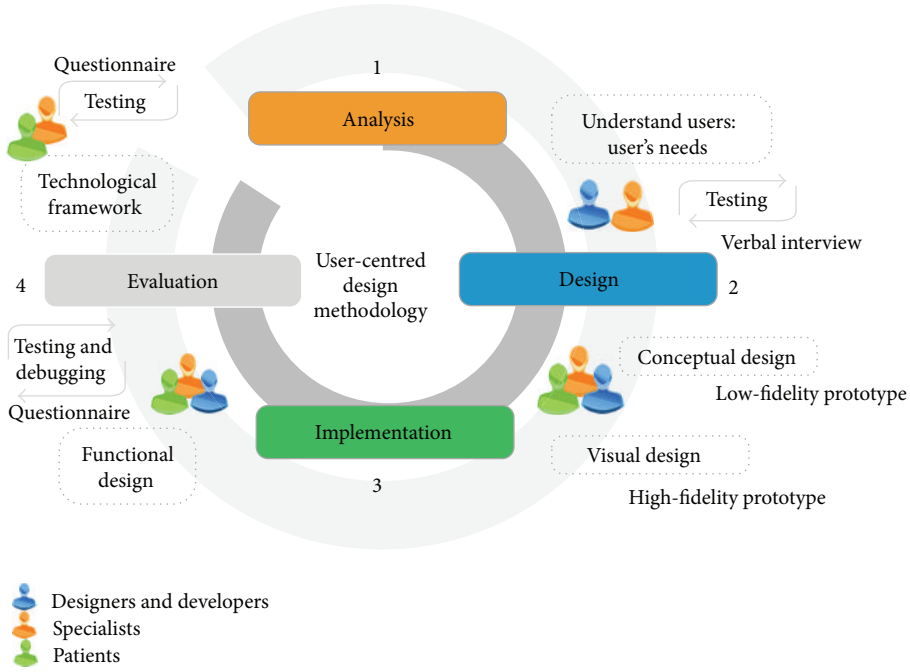


Figure 1: Design process of the telemedicine systems centered on the user. Adaptation of Martínez Alcalá [50].

can be achieved only when the user is actively involved during the design and evaluation of an application or system.

The following sections describe each of the method phases.

(1) Analysis. This phase is focused on three variables: (1) user features, which include cognitive functioning and disabilities, (2) activities characteristics, which include the definition and gathering of content, and (3) definition of the profiles, which includes the identification of end user. (The end user is a user that interacts with a specific device, system, or service. The main users of healthcare ICT are (a) healthcare professionals who work with healthcare information and communication technology applications in hospitals and other healthcare organizations, (b) patients, (c) the general public (regarding the eHealth services), and (d) other supportive parties (e.g., parents, family, and social care workers).)

Then, this phase is focused on each user. The phase begins interviewing patients and specialists to know the information that should be reflected in the system. Then, the objectives that each activity must achieve are identified. Afterwards, the user profiles, which allow the customization of activities and therapy according to the user, are defined. Finally, the analysis and definition of the therapy model are developed.

(2) Design. This phase is focused on the conceptualization of user requirements. This phase begins designing low-fidelity prototypes, which show essential elements of the interface for a specific user. Then, the high-fidelity prototypes, which consist of developing visual design of each end user interface, are developed. Each prototype is assessed by all users to identify any discrepancies or errors during the implementation phase.

(3) Implementation. This phase is focused on creating a layout that includes all the navigation features for each system component (activities, therapies, and progress). The design and implementation phases are iterative. Before implementing a real environment with real users, the prototypes, two things happen: the functional designs are evaluated by all users, and errors that have been identified are corrected and debugged.

(4) Evaluation. This phase is focused on developing the application according to the proposed design and its improvements that resulted from consistent user feedback. The user feedback can be obtained during different stages of the design (conceptual design, visual design, and functional design). In this phase a prototype evaluation with end users is performed.

3.2. MAIA Framework. The MAIA framework is used to support the idea of institutions as a major structure for conceptualizing social systems. The MAIA framework extends and formalizes the components of the IAD4 (an institutional framework that provides a collection of concepts present in a social system with an institutional perspective) to present a metamodel for conceptualizing social systems for agent-based simulation [36].

The MAIA framework has been used in this research by designers with different professional profiles in order to conceptualize the components of telemedicine systems. According to Ferruzca Navarro [37], MAIA is a useful tool to identify those involved in a system. It represents the system through a graphical structure and builds a table to describe its proposed components.

Table 1: Advantages and disadvantages with respect to other systems developed.

Systems	Description of system	Advantage	Disadvantages
The RuralHub Telepsych system [44]	Report on a geriatric telepsychiatry consultation service provided by a tertiary-care hospital to rural nursing homes located up to a few hours' drive away. It has been successfully used with a wide variety of diagnostic groups (such as patients with depression, posttraumatic stress disorder, panic disorder and/or agoraphobia, Alzheimer's disease, schizophrenia, and other mental-health conditions).	(i) Patients show acceptance and adherence to a treatment regimen. (ii) Telepsych is comparable to conventional treatment in outcomes and cost. (iii) Overcomes the traditional therapy (face-to-face sessions), particularly when dealing with patients prone to violence or who are afraid of leaving home for treatment.	(i) It creates "an impersonal atmosphere." (ii) Is problematic for elderly patients with sensory impairments, for treating uncooperative or paranoid patients, and in emergency situation. (iii) Telepsych does not have a collaborative online environment to support exchange of formal and informal information. (iv) There is not involvement of users in the design and development stages of the system.
Portable tele-assessment system [45]	Remote evaluation of the elbow joint with spasticity and contracture in patients with neurological disorders. Especially in patients with spasticity.	(i) Provided physical as well as audiovisual interaction between the clinician and the patient. (ii) Saving time and costs involved in the rehabilitation. (iii) Attractive to the patients since it was designed to be lowcost and portable. (iv) Allows remote monitoring of a remote way the progression of physical treatment without missing the essential part—physical feel.	(i) The spasticity test is limited, because the doctor has to perform therapy exercises, burn for teaching, and the replay to be viewed by the patient. (ii) There is not involvement of users in the design and development stages of the system.
AUBADE system [46]	AUBADE is an integrated platform built for the affective assessment of individuals. The system performs the evaluation of the emotional state.	(i) It has an intelligent emotion recognition module and a facial animation module. (ii) It has databases where the acquired signals along with the subject's animation videos are saved. (iii) Is a multifunctional system that can be utilized in many different ways and in multiple application fields. (iv) The AUBADE system consists of a multisensorial wearable, a data acquisition, and wireless communication module, a feature extraction module.	(i) The system's clinical application is based on the ability of supporting clinical diagnosis related to all the pathologies taking into account if the patient's capability to feel and express emotions is limited or totally absent. (ii) Due to the fact that emotions vary from person to person, the system must be trained by using a variety of subjects and then by testing its performance; this implies an investment of time that often people and specialists do not have.
Telemedical Interventional Monitoring in Heart Failure (TIM-HF) [47, 48]	Wireless Bluetooth system with a personal digital assistant (PDA) that performs automated encrypted transmission via mobile phone of electrocardiogram measurement, blood pressure measurement, and body weight. The telemonitoring system consists, on the one hand, of portable home devices (ECG, and blood pressure measurement and body weight) connected to PDA via a local network. On the other hand, a telemedical workstation with electronic patient record that is a web application with a graphical user interface browser so that incoming measurements generate events according to a set of medical prioritization rules to initiate a workflow-guided review process in the telemedical workstation and its further evaluation by medical professionals.	(i) Prevents hospitalizations by early detection of disease worsening followed by immediate intervention. (ii) Home devices for ECG, body weight, blood pressure, and self-assessment measurement are used. (iii) PDA has a touchscreen option for a scaled self-assessment. (iv) A home emergency call system provides direct contact to health-care professionals. (v) The medical system has been built as an open platform to integrate other home devices for monitoring such as diabetes, chronic obstructive pulmonary disease, anticoagulants, and implantable cardiac device information.	(i) Telemedical centers must operate around the clock every day of year because it requires immediate diagnosis and prompt treatment. (ii) To ensure patient safety, it was required at least that 94% of the total system be available, including the mobile phone network. (iii) Wireless technology is susceptible to faults. Besides there is not defined standards for secure wireless transmission of the data. (iv) The volunteers that proved the system were younger than the anticipated chronic heart population.

Table 1: Continued.

Systems	Description of system	Advantage	Disadvantages
Heart failure case disease management program [49]	Is a care process that verifies the state of a patient illness throughout sending the information concerning his or her vital signs such as weight, systolic blood pressure, heart rate, dyspnoea, asthenia, edema, therapy changes, blood urea nitrogen, creatinine, sodium, potassium, and bilirubine to medical staff in order to support decision-making to prevent haemodynamic imbalance, to reinforce educational support, to optimize therapy, and to improve quality of life and outcomes. The telemedical staff and patients use a touchpad or mobile phone at their home, after having dialed a toll-free number. Then, each parameter was entered in reply to question asked by a recorded voice and a confirmation was requested to each of them.	(i) The program uses a toll-free number. (ii) The overall procedure is managed by an interactive voice response (IVR) system (Appel Electtronica Srl, Turin); therefore, the data transmission did not require operator support. (iii) The daily telemonitoring activities typically began by listening vocal message and taking the appropriate actions. (iv) Optimized therapy and continuous redefinition of the care process. (v) Increases patient's knowledge about management of illness, recognition of initial signs, and symptoms. (vi) The tight relationship between health-care personnel and patients allows coaching in a way that the patient is simulated toward an activate participation in self-management of his or her illness.	Management effectiveness depends on the team management, the intensity of treatment, the parameters monitored, the standardization of managerial algorithms and the characteristics of the patients.
eMental System*	Supports the rehabilitation of the elderly with cognitive impairment through promoting social integration. It provides a cognitive stimulation therapy to the patients, caregivers, and specialists.	(i) It manages automatically the degree of difficulty to suit the cognitive level of each patient. (ii) Provides visual feedback to users' tasks. (iii) The therapy is performed in the comfort (iv) of home. (v) It promotes family to incorporate with patient rehabilitation. (vi) Saving time and costs involved in rehabilitation.	
e-Park system*	The detection of cognitive deterioration of person with Parkinson's disease. By applying the PD-CRS test through Internet.	(i) Standardization and optimization in the application of the PD-CRS test. (ii) Capture and execution of quasi-digital PD-CRS test. (iii) The system automatically manages the patients and the time duration for each session. (iv) Provides visual feedback of the test results.	The first version of the system is not depending on the limitations of the patients.

*Telemedicine system described in this research.

Now, a conceptual model based on the framework for the analysis of interactions between agents (MAIA) is presented [37]. According to Ferruzca Navarro [37], the conceptual model behind this framework is composed of structural components and articulations. The interactions within a system lead to patterns that can be evaluated by the analyst. The MAIA framework views actors as institutional-driven entities. Agents form the key concepts of the modeled system and then are placed within a context.

The main objective of MAIA framework is the identification of the key players involved in a telemedicine system and their relationships.

The main components of MAIA are *organization, subject, artifacts, and context,* which are considered structural agents

within the system that coordinate their actions according to the pursued objective, giving place to other elements such as *object, task activities, products, and representational activity* (see Figure 2).

A fundamental aspect of any system is the analysis and understanding of their composition. In this regard, the components (entities) related to the structure of the telemedicine system are described based on the MAIA model. According to the author in [37], the conceptual model of this methodology consists of structural components and articulation. Each is briefly described here.

3.2.1. Structural Components

(i) Organization. The organization establishes the division of work in the system and their operating rules (procedures). Furthermore, it provides the means of communication and working that subjects have to use in their work. Examples of organization are hospitals, schools, research centers, and laboratories.

(ii) Subject. A subject is an agent who starts a task and is able to interact with other members of the system. The description of a subject can be based on demographics, physical or motor skills and cognitive, emotional, or affective aspects. A subject may perform one or more roles within the system. Examples of subjects include patients, families, caregivers, and specialists.

(iii) Artifact. It is a tangible or intangible resource that allows the execution of a task. The artifacts affect what subjects do and how they do it. There are many ways to understand the artifacts. Besides, there are several ways to describe the artifacts, such as their appearance, use, and personal satisfaction. Examples of artifacts include computers, mobile documents, equipment, and applications.

(iv) Context: Workspace (Physical or Virtual) in Which the subjects and Artifacts Are Developed. The context affects the learning process of individuals as well as their behavior. Examples of spaces include hospital therapy rooms, home, and web applications.

(v) Product: Result of the Interactive Activity between the System Components in Accordance with the Objective Pursued. A device, a service, and a professional are examples of products obtained through various tasks. A "product" could be employed in another system as artifact, procedure, subject, or environment.

3.2.2. Articulation Components

(i) Objective. The organization raises a number of goals that must be achieved. These goals determine the behavior of the subjects. To achieve these goals, people perform tasks which lean on other people, artifacts, and environments.

(ii) Task. The fulfillment of objectives is performed by carrying out a series of tasks that are designed to achieve them. The tasks are assigned based on the role of each person or group and whose complexity can break down into a set of simple activities.

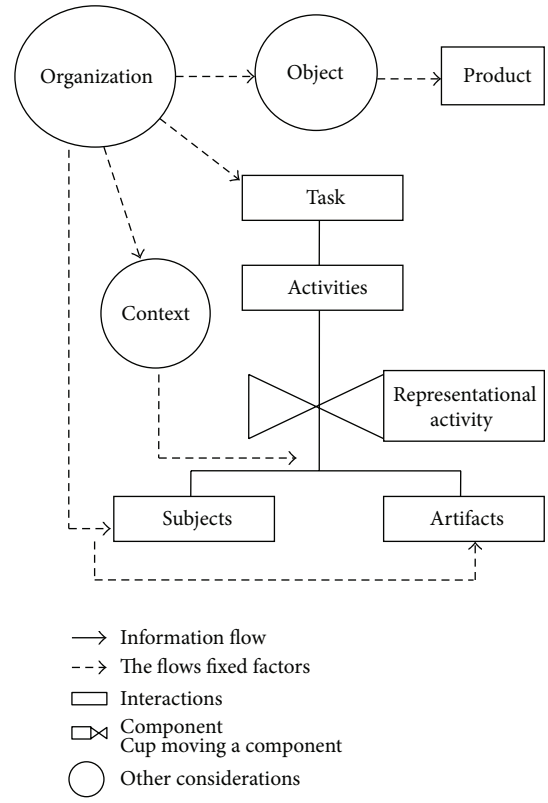


Figure 2: Conceptual model of the telemedicine system components [37].

(iii) Activity: *Set of Steps to Carry Out a Task.* After defining the components of the MAIA framework, the structural components may be immediately identified to make up the system. Similarly, it is possible to define the articulation components. Next, we are showing the questions that guide the identification of system components (see Figure 3).

4. Results

At the Laboratory of Multimedia Applications at the Universitat Politécnica de Catalunya, telemedicine systems have been developed. The telemedicine systems aim to create multimedia tools, available through the Internet, which contribute to improve patients and families life quality. The systems are intended not only to serve in the rehabilitation of patients, but also to therapists and specialists in the process of monitoring the therapy. Next a detailed description of the telemedicine systems created based on the reference model for UCD, described in the previous section, is presented.

4.1. Case Study

4.1.1. Case Study 1: The eMental System. The main objective of the eMental system is to support the rehabilitation of the elderly with cognitive impairment and to promote their

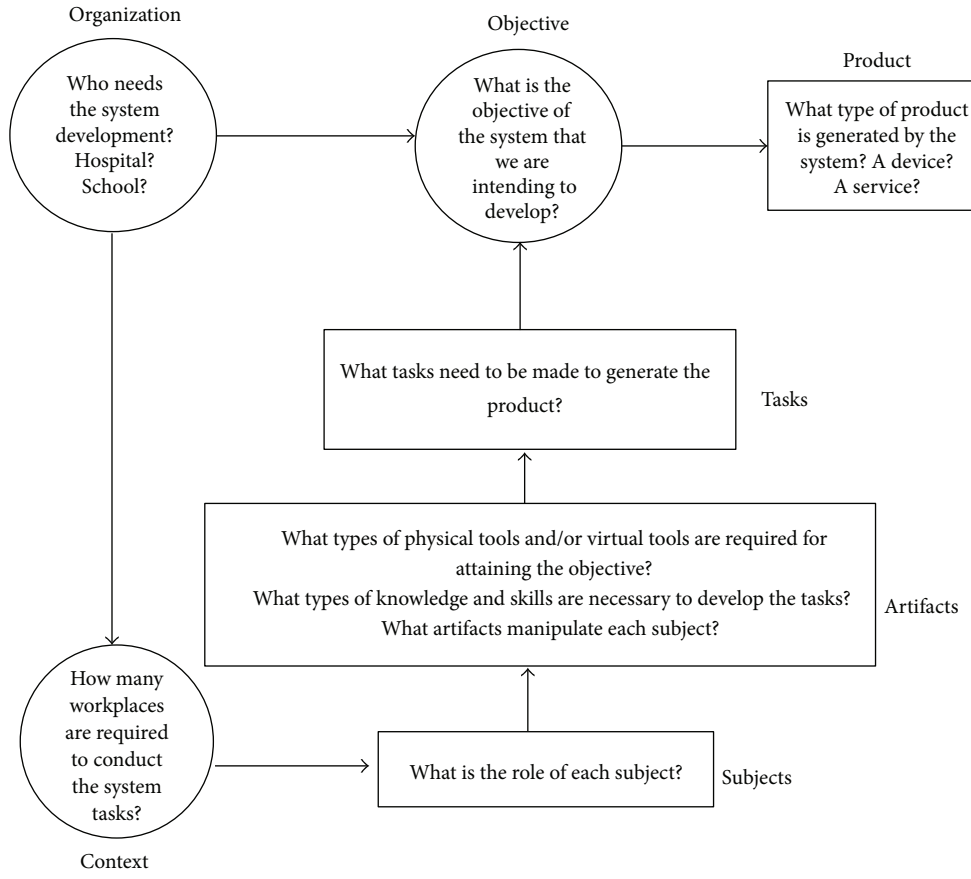


Figure 3: Questions guide to the identification of system components [37].

social integration. The eMental system provides a cognitive stimulation therapy to the patients, caregivers, and specialists. Its primary function is to improve the capabilities of people with any cognitive alteration, through a telemedicine system.

(1) Analysis. The degenerative disorders, which include cortical dementias such as Alzheimer's disease (AD) and subcortical dementias such as Parkinson's disease, are most prevalent. Cognitive impairments also result from other less prevalent conditions as traumatic brain injuries (TBIs); vascular disorders such as strokes; other progressive disorders of the central nervous system such as multiple sclerosis (MS); toxic conditions such as alcoholism; infectious processes such as HIV and AIDS; brain tumors; oxygen deprivation; and metabolic conditions such as diabetes. Cognitive decline also occurs routinely in individuals who are aging "normally" [38]. The patients with cognitive impairment have some physical and mental limitations, so they are sometimes assisted by their caregiver or family to perform daily activities.

As a result of a first approach with a hospital in Spain, it was possible to define the participation of different actors that were involved in the rehabilitation process.

The eMental system is directed to the next users: doctors, patients, caregivers, and other specialists (see Table 2). Next, the approaches obtained from the study are presented as follows:

(i) professional cases (doctors, therapists, and other specialists) require a telemedicine system that helps them monitor and trace patients with cognitive alterations. It also should be able to manage customized therapy sessions;

(ii) patient cases require an asynchronously system that allows patients to access from any point and place without going to the hospital or rehabilitation place. It allows health-care providers to follow the patient's evolution and medication at home and, in general, monitoring the patient from a distance;

(iii) patients suffering from cognitive alterations such as brain injury, mild or moderate Alzheimer, neurodegenerative disorders, psychiatric disorder associated with cognitive impairment. This procedure includes supporting family and/or caregivers that assist in the rehabilitation of patients.

To understand the needs and limitations of each user, a direct observation on the subjects of the research (patients, caregivers, and specialists) was used focusing on their situation, environment, and activity. To do so, verbal interviews were performed to all users interested in research; this allows to collecting their opinions and experiences. Then, these comments were taken into account in the system design process.

Table 2: Descriptions of the user types.

User types	Function
Patient	(i) Requests help the caregiver. (ii) Performs activities assigned by the therapist.
Caregiver	(i) Provides patient instruction. (ii) Encourages the patient. (iii) Gives feedback to the patient.
Therapist	(i) Assigns therapy activities. (ii) Provides indications to the caregiver. (iii) Performs assessment of patients with the results obtained in each of the assigned activities. (iv) Facilitates feedback and helps the patient.
Doctor	(i) Provides indications to the therapist. (ii) Provides medical evaluation of patient.
Administrator	Manages system resources

After understanding the users' goals and needs, the next step consisted in organizing the information according to the MAIA framework [37]. In order to facilitate the understanding of the main actors involved in the therapeutic process and to allow the development of further systems adapted to the actual needs and circumstances, the representation of the therapy process for the eMental system by following MAIA methodology is shown (see Figure 4).

(2) Design. The eMental system is designed for specialists in cognitive therapy. At the design and development process, different roles were identified as follows: (a) the medical team defined the contents; (b) the design team proposed the graphical user interface (GUI) (*Graphic User Interface* represents the information and actions available to a user through icons and graphical elements [39]) that attempted to represent communication between the patient and the specialist; and finally (c) the development team was responsible for programming and executing the contents of the system. The eMental system considered important capabilities such as attention and concentration, executive functions (reasoning, planning), perception and knowledge, language, and computation special orientation.

In order to characterize the contents of the system as far as possible, a conceptual design of the eMental system was schematized after conducting a literature review. This process was conducted with the support of several medical specialists and taking into account the opinions and experiences gathered in the direct observation. In the traditional therapy, the therapist performs the therapy through using pencil and paper supported with a picture book with cognitive stimulation exercises.

The eMental system was designed in a way that the therapy can be performed by the patient in the comfort of his or her home. Each of the cognitive stimulation exercises mentioned in the picture book were adapted and customized in the system so that patients could perform their exercises in a more comfortably and easily way. Furthermore, the therapists could perform the rehabilitation therapy more quickly and with a greater control over the results. This allows to attain a reduction of costs and time for both the patient and the hospital. To achieve this, the therapist classified each of the years in a total of 6 items, while the design team digitized each of the necessary fonts (*images, photographs, diagrams, etc.)* to design the exercises within the system.

Once the components and contents were defined, the team performed several low-fidelity prototypes (sketches) through designing quick and easy user interface. These prototypes were presented to the medical team and the development team in order to get comments and new ideas regarding the design of the system framework.

Therefore, a better designed high-fidelity prototype was obtained. Next, three important areas of the system are determined in order to understand the tasks, the therapy, and the progress. These are described as follows:

(i) the task area contains customized activities that the doctor or therapist considers relevant to the patient's rehabilitation;

(ii) the therapy area has n modules that integrate memory exercises, numbers, letters, and drawings. It also has three levels which are adjusted to the patient's evaluation;

(iii) the progress area: the results obtained during the therapy process can be displayed graphically to the users.

This eMental system consists of a series of 96 exercises such as calculation, memory, attention, orientation, language, and visuospatial exercises designed to address mental capacities like attention and concentration, executive functions (reasoning, planning), perception and knowledge, language, calculation, and special orientation. Next, some prototypes designed for representing patient tasks are shown.

The eMental system provides users a greater possibility of easily access to the technology world and, at the same time, strengthens their mental functions with exercises that are displayed in a web environment. In addition, the system has an asynchronous mode, meaning that the patient performs exercises with the assistance of a family member or caregiver, anytime, anywhere. Furthermore, the system records the patient's activities to be evaluated by a doctor or therapist and the progress of the therapy is recorded visually.

As a result, the first design proposals obtained have an intuitive graphical interface that facilitates the interaction between the user and the system without having advanced computer skills. Moreover, each of the designed activities increases the self-esteem and reinforces the skills the patient still preserves, reducing frustration towards the therapy through encouraging messages to the successes and failures of the patient. This leads to improved results, such as increasing the attention of users and minimizing external distraction.

(3) Implementation. The implementation of the first version of the eMental system involved a hospital in the south of Spain. The hospital provided a rehabilitation team which consisted in *a medic, a physiotherapist, a patient, and a caregiver* and also furnished with areas for therapy rooms.

The multimedia engineering laboratory team was responsible for installing the technological artifacts (touchscreen,

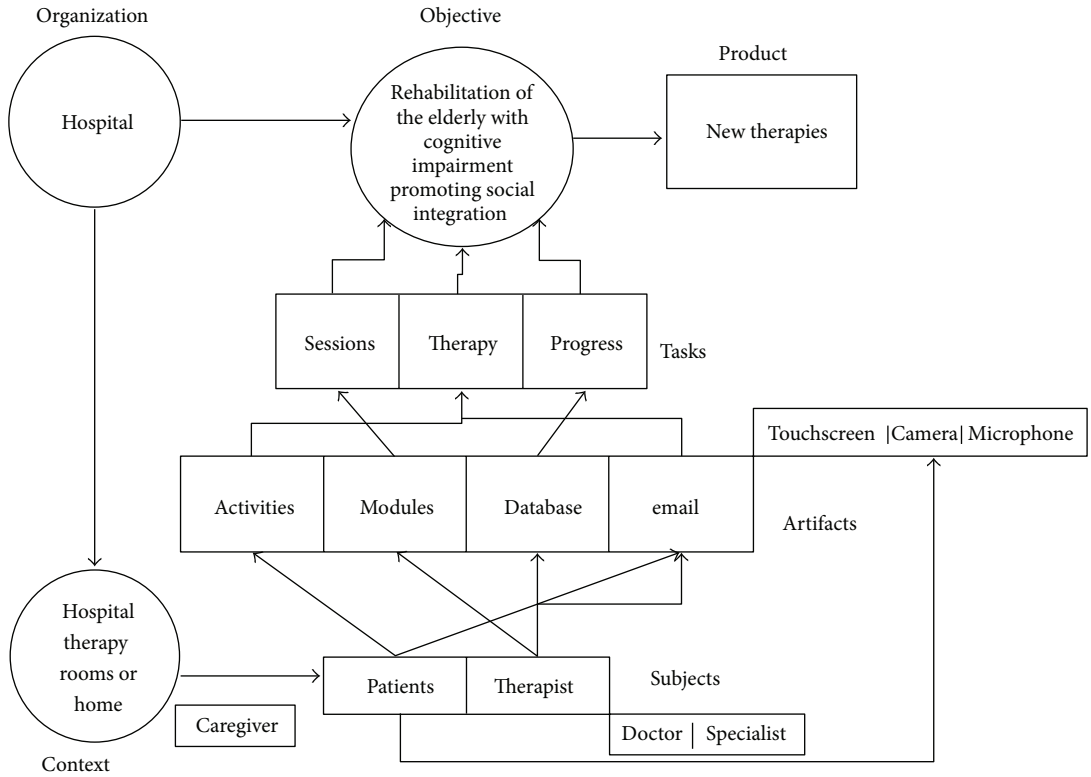


Figure 4: Representation of the therapy process for the eMental system by following MAIA methodology.

camera, and microphone) within the therapy rooms. Besides, they presented the proposal for the eMental system design by considering a detailed visualization of the tasks each patient should perform and the visualization process of the therapy each therapist should receive.

In order to evaluate the first design proposals, there were selected users who had the next features: (1) were familiar with cognitive impairment (such as doctors, therapists, patients, caregivers, and other specialists); (2) had basic computer skills, such as sending and receiving e-mail and capturing images from a camera or electronic device; (3) had knowledge of the test application, time availability, and exchange experiences on the issue; and finally (4) were interested in the rehabilitation through ICTs. The characteristics of the study group will be described in the following section.

The next step to the system is conducting tests with bigger groups, considering variables such as acceptance and usability.

(4) Evaluation. The researchers selected a group of potential users to analyze and evaluate the functional model designed from the prototypes improvements. A total of nine users attended the first presentation of functional layout of the system (two patients, two caregivers, one therapist, one doctor, two engineers, and one web designer).

To evaluate the usability of the system, first, the graphical interface and digitized exercises were presented. Later, the use and functionality of the system were explained in detail. Finally, verbal questions were performed to obtain each present person's point of view. As a result, on the one hand, small changes in the graphical interface (color, font size) were indicated. On the other hand, difficulty levels to the exercises were added.

At present, in order to improve the opportunity areas that have been identified in the first evaluation feedback, the system is in a redesign phase.

4.1.2. Case Study 2: The e-Park System. The main objective of the e-Park system is the detection of cognitive deterioration of a person with Parkinson's disease. This is achieved through a telemedicine system that allows evaluating patients with a disease scale of PD-CRS by using telemedicine system.

(1) Analysis. Parkinson's disease has degenerative neurological symptoms that initially affect the motor system; however, it is evidenced that other areas could be affected by the progression of the disease, such as the cognitive and autonomic systems. The cognitive affectation is concentrated in the executive functions area, visuospatial skills and some forms of memory and language [40, 41]. The main symptoms of this disease are poor control of movements: shaking, sluggishness, stiffness, and abnormalities of posture and walking. Moreover, the affected person requires different cares to help him/her perform daily activities. The assistance of family members or caregivers for improving his/her life quality is required.

As a result of a first approach with a hospital in Spain, it was possible to define the participation of the different actors that were involved in a rehabilitation process.

The e-Park system is focused on the next users: doctors, patients, caregivers, and other specialists (as described in

Section 4.1.1, Table 2). Next, the approaches obtained from the study are presented as follows:

(i) professional cases (doctors, therapists, and others specialists) require a telemedicine system that helps them monitor and trace patients with cognitive deterioration. It also should be able to manage a customized therapy sessions;

(ii) patient cases require an asynchronously system that allows access to patient from any point and place without going to the hospital or rehabilitation place;

(iii) its require system to allow remote patient monitoring and the evolution and implementation of medication is administered at home;

(iv) the patients are those who have Parkinson's disease. This procedure includes supporting family and/or caregivers that assist in the rehabilitation of patients.

Same as the previous case study, to understand the needs and limitations of each user, a direct observation was used to understand their situation, environment, and activity. Then, all comments were taken into account in the system design process.

After understanding the goals and needs of users, the next step consisted in organizing the information according to the MAIA framework [37]. In order to facilitate the understanding of the main actors involved in the therapeutic process and to allow further system development adapted to the actual needs and circumstances, the representation of the therapy process for the e-Park system by following MAIA methodology is shown (see Figure 5).

(2) Design. The e-Park system is planned by expert doctors experienced in specialized drugs treatment (the selection of drug treatment/therapy should be made after the patient with Parkinson's disease has been properly informed of drug efficacy [42]) for rehabilitation of Parkinson's disease. The e-Park system is designed for specialists in the application of the PD-CRS test. At the design and development stages, different roles were identified as follows: (a) the medical team defined the activities for the test; (b) the design team proposed the graphical user interface that attempted to represent communication between patient and specialist; and finally (c) the development team was responsible for programming and executing the contents of the system.

In order to characterize the contents of the system as far as possible, a conceptual design of the e-Park system was schematized after conducting a literature review. This process was conducted with the support of several medical specialists and taking into account the opinions and experiences gathered in the direct observation. In the traditional therapy, the therapist performs the therapy through using pencil and paper, supported with a test called *Cognitive Rating Scale (PD-CRS).*

The e-Park system was designed in a way that the therapy can be performed by the patient in the comfort of his or her home. Each question on the test was adapted and customized within the system so that the therapist could easily record the results of their patients, achieving a reduction of costs and time for both the patient and the hospital. Besides, the patients could perform the exercises in a more comfortable and easily way.

The test (PD-CRS) includes nine divisions with a maximum total score of 92 as follows:

(i) subcortical functions: attention (10), short memory (10), working memory (10) and delayed memory (12), verbal fluency and alternating action, and spontaneous drawing of clock (10);

(ii) cortical functions: designation (20) and a clock copy (10).

The verbal fluency and alternating do not have maximum scores. They are obtained by adding the rating for subcortical subscales, the cortical and the total PD-CRS. Better punctuation provides better cognitive level.

Once the components and contents were defined, the team performed several low-fidelity prototypes (sketches) through designing quick and easy user interface. These prototypes were presented to the medical team and the development team in order to get comments and new ideas regarding the design of the system framework.

Therefore, a better designed high-fidelity prototype was obtained. Next, two application modes of the system are determined (see Figure 6):

(1) online assistance, which means helping the specialist perform a real-time evaluation of the patient;

(2) self-evaluation, meaning the evaluation of the patient himself or herself with this assistance of his or her family members and caregivers previously trained in this activity.

The e-Park system has an asynchronous mode (remote real-time communication), meaning that the patient performs the test exercises with the assistance of a family member or caregiver, anytime and anywhere. Furthermore, it allows the specialist to manage the application of the system and their patients through using a web application such as video conferencing (see Figure 7).

The patient performs the test in front of a screen by following the specialist's instructions, and the specialist interacts with another application. This allows guiding the execution of the PD-CRS test and the record of the results of the patient.

(3) Implementation. The implementation of the first version of the e-Park System involved two hospitals in southern Spain. The hospitals provide a rehabilitation team, which consisted in a *medic, a therapist, a neurologist, a patient, and a caregiver* and also furnished with areas for therapy rooms. The multimedia engineering laboratory team was responsible for installing the technological artifacts (touch screen, camera, and microphone) within the therapy rooms. Besides, they presented the proposal of functional design by considering a detailed visualization of the test PD-CRS.

In order to evaluate the first design proposals, there were selected users who had the next features: (1) were familiar with Parkinson's disease (such as doctors, therapists, patients, caregivers, and other specialists); (2) had basic computer

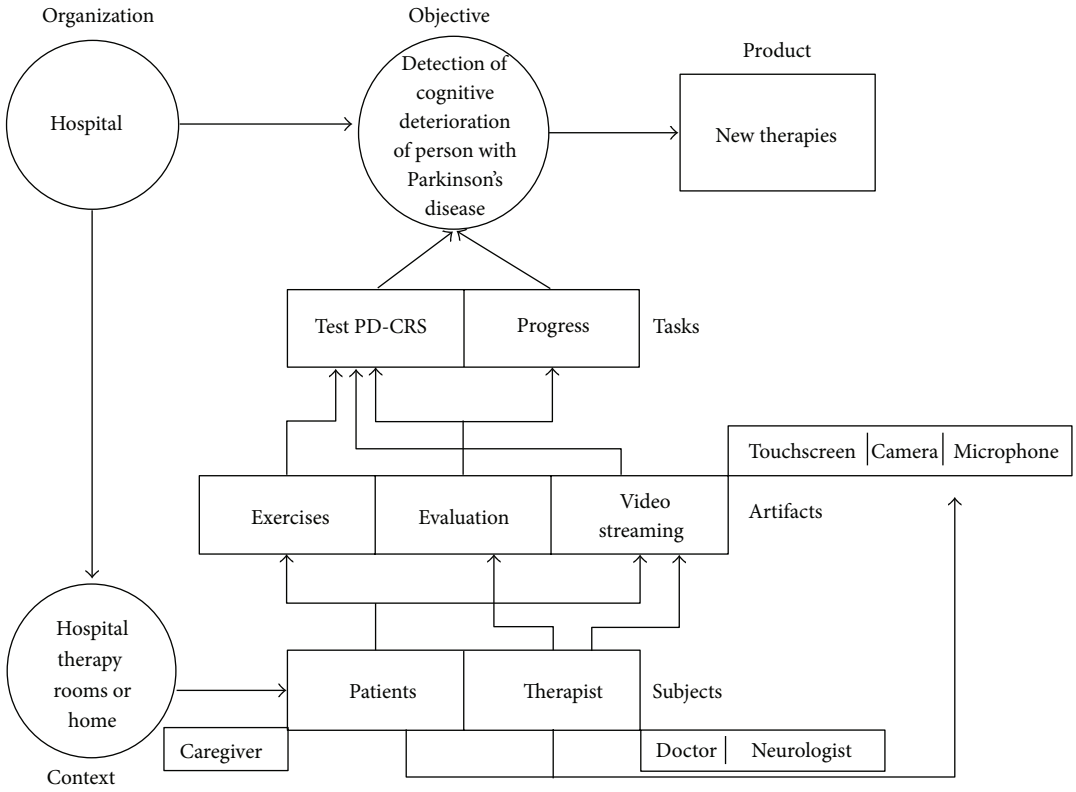


Figure 5: Representation of the therapy process for the e-Park system by following MAIA methodology.

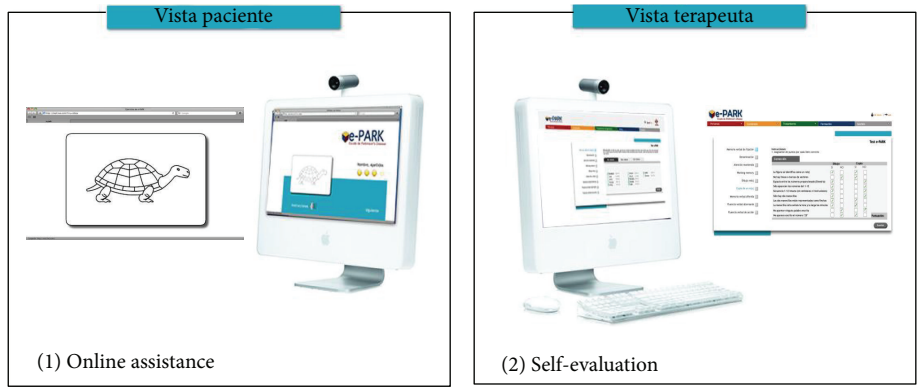


Figure 6: System application modes.

skills, such as sending and receiving email and capturing images from a camera or electronic device; (3) had knowledge of the test application, time availability, and exchange experiences of the issue; and finally (4) were interested in the rehabilitation through ICTs. The characteristics of the study group will be described in the following section.

The next step to the system is conducting tests with bigger groups, considering variables such as acceptance and usability.

(4) Evaluation. It is important to mention that eMental and e-Park systems are in a phase of evaluation and redesign. Because of this, this research shows the first results obtained in the usability evaluation of the two systems.

Qualitative evaluation is used in the research. Therefore, the number of participants in the case study is small. Furthermore, the qualitative analysis allows the data to be manipulated in order to extract a relevant significance regarding the objective of the research.

The patients were selected for a consecutive period of 12 weeks between April 2010 and July 2010. Appropriate cases were defined to those following patients:

(i) patient 1: patient with primary Parkinson's disease;

(ii) patient 2: patient with primary Parkinson's disease;

(iii) patient 3: patient in the predementia stage of Alzheimer's disease;

(iv) patient 4: patient in the early stage of Alzheimer's disease.

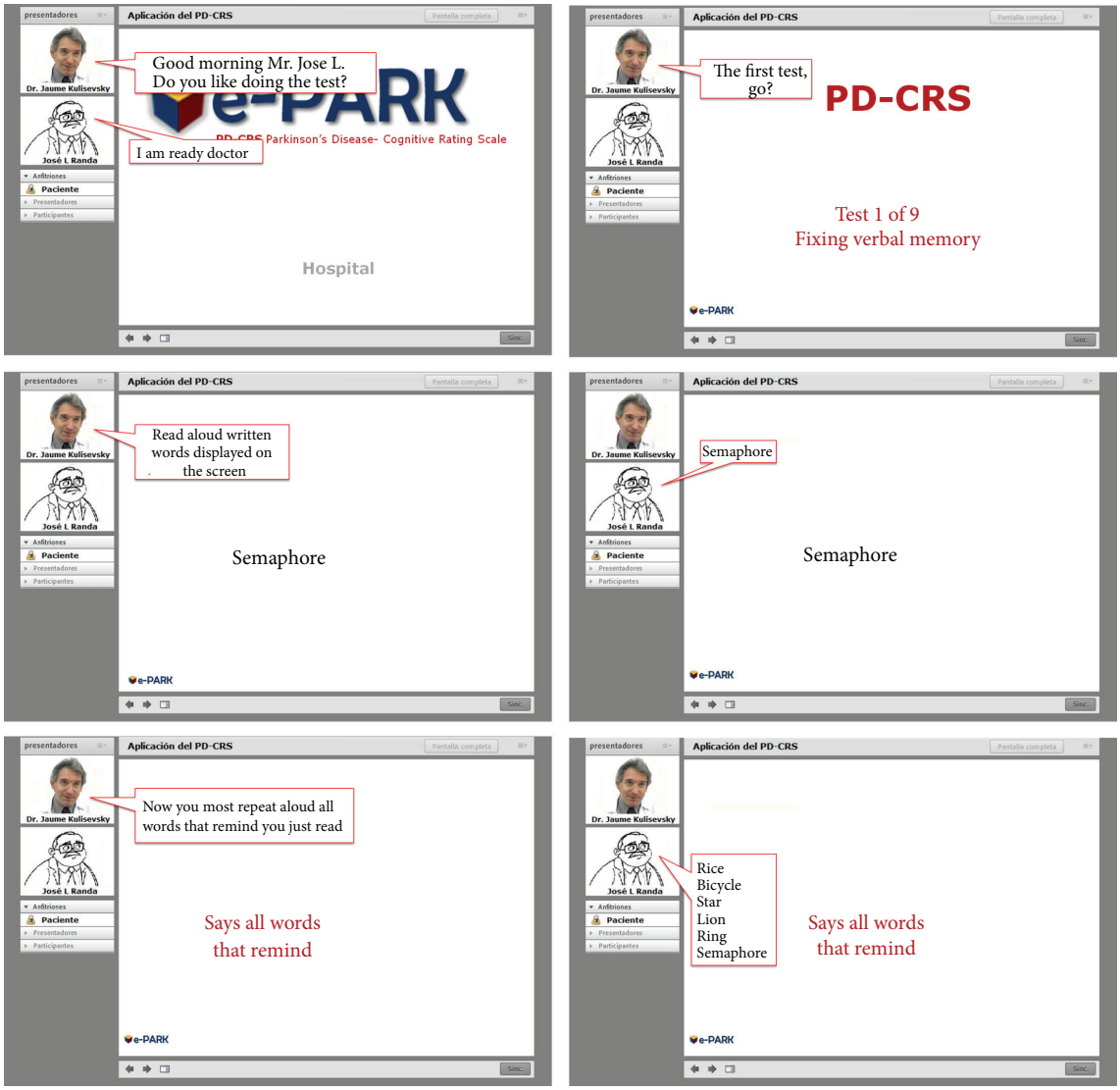


Figure 7: Sequence of the therapy.

The age range for this study is 60 years old, with only one case being under the age of 50 years. The rate for men is 75% and for women 25%.

The work team showed the functional prototype of the eMental system. The doctors and therapists visualized the tool by running the system and exercises of the patients.

Moreover, the work team showed the functional prototype of the e-Park system. The patient performed the test in front of a screen by following the specialist's instructions, and the specialist interacted through another application.

Finally, the complete evaluation of the first proposal of the two systems took place on February 24, 2011. It involved a total of 17 people, including four patients, four caregivers, two therapists, one neurologist, three doctor, two engineers, and one web designer.

In order to carry out a first evaluation of the two systems, a questionnaire based on parameters of usability and technology acceptance (the technology acceptance evaluates a series of factors that influence the decision on *how* and *when* the patient will use the technology [43]) was applied. The questionnaire composed of 16 closed questions was answered by the studying group. The mean time of the evaluations duration was approximately 30 minutes.

The questionnaire applied to the study group in order to get information about parameters of usability and acceptance of technology is shown in Table 3.

As shown previously, the questionnaire is composed of 15 closed questions, grouped into five parameters: easiness of navigation, learnability, satisfaction, operability, and functionality. The questionnaire also included one open question in order to obtain different comments and feedback of the system.

The scale of measure used in the questionnaire corresponds to a Likert scale of seven points. The options in the scale was as follows: (a) the range of values between 1 and 3 were selected for expressing disagreement; (b) the value 1 was selected for strong disagreement; (c) the value 4 was selected for a neutral opinion; and finally (d) the range of values

Table 3: Questions applied to the study group. Adaptation of Martínez Alcalá, 2012 [50].

Parameter	Questions
Easiness of navigation	(i) Is my interaction with telemedicine system clear and understandable? (ii) Was it easy for me to use the telemedicine system? (iii) Do I find easy to use the telemedicine system?
Learnability	(i) Is it easy for me to learn how to operate the telemedicine system? (ii) Is the telemedicine system design friendly? (iii) Are the instructions given in therapy clear and easy to understand?
Satisfaction	(i) Is it helpful to implement a telemedicine system for rehabilitation? (ii) Is the telemedicine system design an attractive idea? (iii) Does the telemedicine system provide an attractive rehabilitation environment? (iv) Do I like working with telemedicine system?
Operability	(i) Does the telemedicine system design meet your expectations? (ii) Does the design (appearance, color, shape, etc.) clearly show that it is a support system to improve the cognitive impairment problem? (iii) Is the telemedicine system capable of correctly representing improvements experienced by the treatment? (iv) Do I find the telemedicine system useful in my rehabilitation?
Functionality	Does the design of therapy allow its activities easier?
Open question	In general, do you think that undergoing a rehabilitation process through the telemedicine system increases the chances of improving your current condition?

between 5 and 7 was selected for expressing agreement, where 7 meant totally agree.

Then, a total of 16 validated questionnaires were obtained. The rate of response of the questionnaires was 82.35% (referred to the total percentage of medical personnel and patients in the hospital). The summary of results obtained from the application of the questionnaire is shown below (see Figure 8).

In the analysis of the first parameter, *ease of navigation*, the users showed a positive agreement (7), asserting that telemedicine systems designed are of easy navigation, and therefore, the patients will be able to use it at home. A positive value in the first parameter finally promotes the implementation of telemedicine systems for the rehabilitation of patients with neurological diseases.

In the second parameter, *learnability*, the users showed a positive agreement (7), reflecting to agree on the design of telemedicine systems is friendly, clear, and easy to understand. This indicates the systems designed in this telemedicine research taking into consideration the needs of the hospital and patients.

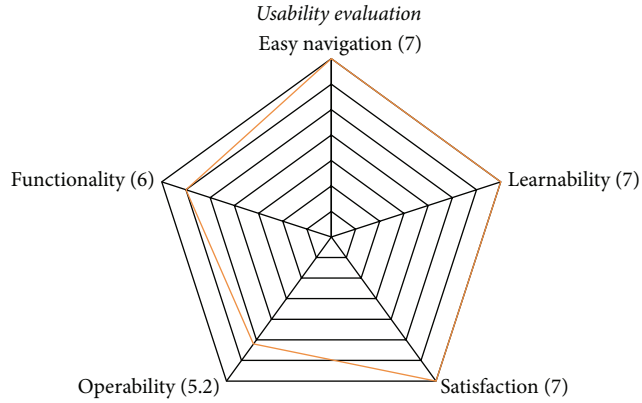


Figure 8: The results obtained of usability evaluation.

The third analyzed parameter is *satisfaction;* the user showed a favorable result (7), manifesting a positive agreement that telemedicine systems provide an attractive environment for rehabilitation.

In the parameter of *operability*, there is a very minimal margin between the answers provided (5.2), which indicates that in the future, the users need more time to evaluate the usability of the system.

Concerning the analysis of the last parameter, functionality, the users showed a positive agreement (6), affirming that the design of therapy allowed the patients to perform rehabilitation exercises in a more easy and attractive way.

4.2. Technological Attributes. Table 4 shows a comparative analysis according to the technological attributes that are provided by the telemedicine systems developed in this research.

This comparison shows that the systems described in this research are designed step by step, with the purpose of customizing therapies according to the profile and disability of patients, taking into account the needs of patients throughout the system design process.

5. Conclusions

Telemedicine is gradually, if not rapidly, becoming a technological and clinical reality. Therefore, it is essential to address the challenge that exists in the successful evaluation of a telemedicine system. By reviewing the literature related to telemedicine systems, we noticed the necessity to concentrate on the specific user requirements, particularly referring to patients, in order to develop an intuitive and effective system.

This research work proposes a framework based on the UCD methodology and MAIA framework, to design systems customized to particular users with specific characteristics and needs, increasing the acceptance and satisfaction in the users. An important feature of the proposed framework is the involvement of end users through the design processes, which allows collecting important data of user needs. Moreover, the reliability and validation of the first functional layouts are increased. Therefore, this framework becomes the source to optimizing the design of the telemedicine systems interfaces

Table 4: Comparison of systems technological attributes.

	Devices	Compatibility	Complexity	Accessibility	Portability	Satisfaction and acceptance
(1) The RuralHub Telepsych system	E-mail, fax, and telephone	X	X	—	—	X
(2) Portable teleassessment system	Cameras, microphones, and PC	X	X	—	—	—
(3) AUBADE system	Integrated platform and PC	X	—	—	X	X
(4) Telemedical interventional monitoring in heart failure	Personal digital assistant, home devices (ECG), and Web application	—	X	—	X	X
(5) Heart failure case disease management program	Touchpad or mobile phone	X	—	X	X	X
(6) eMental system	Web application, cameras, microphone, and touchscreen	X	X	X	X	X
(7) e-Park System	Web application, cameras, microphone, and touchscreen.	X	X	X	X	X

according to users' real needs. The proposed framework is composed of four phases: analysis, design, implementation, and evaluation.

Besides, this paper shows the performance of the proposed framework through implementing it in two case studies to design two systems: eMental system and e-Park system. These telemedicine systems were developed in real hospital environments in order to focus on monitoring the rehabilitation of patients with neurological disorders. All designed telemedicine systems were customized to cognitive and physical limitations of the patients. This way, they provided visual information of patient progress to the medical staff.

As main benefits, this approach reduces significantly the time of patients in hospitals, without reducing the continuous monitoring of patients. Moreover, it facilitates the flexible interaction between patient and doctor through using web application. This shows the efficiency of using the proposed framework to design telemedicine systems.

After implementing the proposed framework to design the two systems, we can conclude that, to achieve improvements in quality health care and to reduce errors, researchers and system developers must work together to integrate the knowledge of user-centered design toward the design of new systems customized to users with specific needs. Moreover, there is a need of collaboration among medical team, the design team, and the development team to ensure a well design and perform of telemedicine systems.

5.1. Future Work. This work proposes the adoption of user-centered design (UCD) methodology for designing and developing telemedicine systems in order to support the rehabilitation of patients with mental deficiencies. Then, four research lines are identified from this research:

(1) deploy the proposed framework in other telemedicine systems and include other related technology, in order to identify more findings and get more favorable results;

(2) develop tailored versions of telemedicine system for mobile devices;

(3) implement the proposed approach in the treatment and rehabilitation therapy file;

(4) incorporate intelligent agents to support the patient and medical staff in telemedicine systems.

References

[1] F. Shiferaw and M. Zolfo, "The role of information communication technology (ICT) towards universal health coverage: the first steps of a telemedicine project in Ethiopia," *Glob Health Action*, vol. 5, pp. 1–8, 2012.

[2] J. Li, J. Westbrook, J. Callen, and A. Georgiou, "The role of ICT in supporting disruptive innovation: a multi-site qualitative study of nurse practitioners in emergency departments," *BMC Medical Informatics & Decision Making*, vol. 12, article 27, 2012.

[3] L. Teixeira, C. Ferreira, and B. S. Santos, "User-centered requirements engineering in health information systems: a study in the hemophilia field," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 160–174, 2012.

[4] M. Borowski, S. Siebig, C. Wrede, and M. Imhoff, "Reducing false alarms of intensive care online-monitoring systems: an evaluation of two signal extraction algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2011, Article ID 143480, 11 pages, 2011.

[5] M. Rusu, G. Saplacan, G. Sebestyen, N. Todor, L. Krucz, and C. Lelutiu, "eHealth: towards a healthcare service-oriented boundary- less infrastructure," *International Journal of Telemedicine and Applications*, vol. 27, no. 3, pp. 1–14, 2010.

[6] C. Galán-Retamal, R. Garrido-Fernández, S. Fernández-Espínola, and V. Padilla-Marín, "Monitoring polymedicated elderly patients in a health care unit," *Farmacia Hospitalaria*, vol. 34, no. 6, pp. 265–270, 2010.

[7] R. Wootton, A. Geissbuhler, K. Jethwani et al., "Long-running telemedicine networks delivering humanitarian services: experience, performance and scientific output," *Bulletin of the World Health Organization*, vol. 90, no. 5, pp. 321–400, 2012.

[8] V. A. Wade, J. Karnon, A. G. Elshaug, and J. E. Hiller, "A systematic review of economic analyses of telehealth services using real time video communication," *BMC Health Services Research*, vol. 10, article 233, 2010.

[9] M. Kurki, H. Hätönen, M. Koivunen, M. Anttila, and M. Välimäki, "Integration of computer and Internet-based programmes into psychiatric out-patient care of adolescents with depression," *Informatics for Health and Social Care*, vol. 37, no. 3, pp. 1–11, 2012.

[10] C. D. Melas, L. A. Zampetakis, A. Dimopoulou, and V. Moustakis, "Modeling the acceptance of clinical information systems among hospital medical staff: an extended TAM model," *Journal of Biomedical Informatics*, vol. 44, no. 4, pp. 553–564, 2011.

[11] C. Lin, I. C. Lin, and J. Roan, "Barriers to physicians' adoption of healthcare information technology: an empirical Study on multiple hospitals," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1965–1977, 2012.

[12] R. Bernard, S. McNeil, D. Cook, K. Agarwal, and G. R. Singhal, "Preparing for the changing role of instructional technologies in medical education," *Academic Medicine*, vol. 86, no. 4, pp. 435–439, 2011.

[13] N. L. Ruxwana, M. E. Herselman, and D. P. Conradie, "ICT applications as e-health solutions in rural healthcare in the Eastern Cape Province of South Africa," *Health Information Management Journal*, vol. 39, no. 1, pp. 17–26, 2010.

[14] N. Ruxwana, *Technology Assessment of Rural Hospitals in the Eastern Cape Province: Knowledge, Adoption, Access, and Availability of E-Health Solutions for Improved Health Care Services Delivery in Rural Hospitals*, VDM Dr. Müller, 2009.

[15] N. J. Borges, A. M. Navarro, A. Grover, and J. D. Hoban, "How, when, and why do physicians choose careers in academic medicine? A literature review," *Academic Medicine*, vol. 85, no. 4, pp. 680–686, 2010.

[16] D. Veneri, "The role and effectiveness of computer-assisted learning in physical therapy education: a systematic review," *Physiotherapy Theory and Practice*, vol. 27, no. 4, pp. 287–298, 2011.

[17] F. H. Barneveld Binkhuysen and E. R. Ranschaer, "Teleradiology: evolution and concepts," *European Journal of Radiology*, vol. 78, no. 2, pp. 205–209, 2011.

[18] A. D. Sorknæs, H. Madsen, J. Hallas, P. Jest, and M. Hansen-Nord, "Nurse tele-consultations with discharged COPD patients reduce early readmissions—an interventional study," *Clinical Respiratory Journal*, vol. 5, no. 1, pp. 26–34, 2011.

[19] P. Pawar, V. Jones, B. J. F. van Beijnum, and H. Hermens, "A framework for the comparison of mobile patient monitoring systems," *Journal of Biomedical Informatics*, vol. 45, no. 3, pp. 544–556, 2012.

[20] S. C. Peirce, A. R. Hardisty, A. D. Preece, and G. Elwyn, "Designing and implementing tele- monitoring for early detection of deterioration in chronic disease: defining the requirements," *Health Informatics Journal*, vol. 17, no. 3, pp. 173–190, 2011.

[21] C. I. Reis, C. S. Freire, J. Fernández, and J. M. Monguet, "Patient centered design: challenges and lessons learned from working with health professionals and schizophrenic patients in e-therapy contexts," *Communications in Computer and Information Science*, vol. 221, no. 1, pp. 1–10, 2011.

[22] M. B. Ayed, H. Ltifi, C. Kolski, and A. M. Alimi, "A user-centered approach for the design and implementation of KDD-based DSS: a case study in the healthcare domain," *Decision Support Systems*, vol. 50, no. 1, pp. 64–78, 2010.

[23] W. Palmas, S. Shea, J. Starren et al., "Medicare payments, healthcare service use, and telemedicine implementation costs in a randomized trial comparing telemedicine case management with usual care in medically underserved participants with diabetes mellitus (IDEATel)," *Journal of the American Medical Informatics Association*, vol. 17, no. 2, pp. 196–202, 2010.

[24] M. Aanesen, "Smarter elder care? A cost-effectiveness analysis of implementing technology in elder care," *Health Informatics Journal*, vol. 17, no. 3, pp. 161–172, 2011.

[25] T. N. Butler and P. Yellowlees, "Cost analysis of store-and-forward telepsychiatry as a consultation model for primary care," *Telemedicine Journal and E-Health*, vol. 18, no. 1, pp. 74–77, 2012.

[26] D. Wallach and S. C. Scholz, "User-centered design: why and how to put users first in software development," *Software for People Management for Professionals*, vol. 1, pp. 11–38, 2012.

[27] S. Gardiner and T. Hartzell, "Telemedicine and plastic surgery: a review of its applications, limitations and legal pitfalls," *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 65, no. 3, pp. e47–e53, 2012.

[28] P. K. Raju and S. Prasad, "Telemedicine and cardiology decade of our experience," *Journal of Indian College of Cardiology*, vol. 2, no. 1, pp. 4–16, 2012.

[29] A. B. Bynum and C. A. Irwin, "Evaluation of the effect of consultant characteristics on telemedicine diagnosis and treatment," *International Journal of Telemedicine and Applications*, vol. 2011, Article ID 701089, 8 pages, 2011.

[30] B. Fong, A. C. M. Fong, and C. K. Li, *Telemedicine Technologies: Information Technologies in Medicine and Telehealth*, Wiley-Blackwell, 2010.

[31] A. While and G. Dewsbury, "Nursing and information and communication technology (ICT): a discussion of trends and future directions," *International Journal of Nursing Studies*, vol. 48, no. 10, pp. 1302–1310, 2011.

[32] E. L, "Telemedicine: the future of outpatient therapy?" *Clinical Infectious Diseases*, vol. 51, no. S2, pp. S224–S230, 2010.

[33] H. J. Rogove, D. McArthur, B. M. Demaerschalk, and P. M. Vespa, "Barriers to telemedicine: survey of current users in acute care units," *Telemedicine and E-Health*, vol. 18, no. 1, pp. 48–53, 2012.

[34] Y. Kowitlawakul, "The technology acceptance model: predicting nurses' intention to use telemedicine technology (eICU)," *Computers Informatics Nursing*, vol. 29, no. 7, pp. 411–418, 2011.

[35] A. G. Ekeland, A. Bowes, and S. Flottorp, "Effectiveness of telemedicine: a systematic review of reviews," *International Journal of Medical Informatics*, vol. 79, no. 11, pp. 736–771, 2010.

[36] A. Ghorbani, V. Dignum, and G. Dijkema, "An analysis and design framework for agent-based social simulation," in *Advanced Agent Technology*, vol. 7068 of *Lecture Notes in Computer Science*, pp. 96–112, 2012.

[37] M. V. Ferruzca Navarro, *Estudio teórico y evidencia empírica de la aplicación del marco teórico de "Cognición Distribuida" en la gestión de sistemas de formación e- Learning [Tesis Doctoral]*, Universidad Politécnica de Cataluña (UPC), 2008.

[38] E. Cole, "Patient-centered design: interface personalization for individuals with brain injury," in *Proceedings of the 6th international conference on Universal access in human-computer interaction: users diversity (UAHCI '11)*, pp. 291–300, 2011.

[39] D. Pavliha, M. Reberšek, and D. Miklavčič, "A graphical user-interface controller for the biomedical high-voltage signal generator," *Journal of Electrical Engineering and Computer Science*, vol. 78, pp. 79–84, 2011.

[40] J. M. Shulman, P. Jager, and M. B. Feany, "Parkinson's disease: genetics and pathogenesis," *Annual Review of Pathology*, vol. 6, pp. 193–222, 2011.

[41] D. Weintraub, J. Koester, M. N. Potenza et al., "Impulse control disorders in Parkinson disease: a cross-sectional study of 3090 patients," *Archives of Neurology*, vol. 67, no. 5, pp. 589–595, 2010.

[42] E. Unlu, A. Cevikol, B. Bal, E. Gonen, O. Celik, and G. Kose, "Multilevel botulinum toxin type a as a treatment for spasticity in children with cerebral palsy: a retrospective study," *Clinical Science*, vol. 65, no. 6, pp. 613–619, 2010.

[43] B. Šumak, M. Heričko, M. Pušnik, and G. Polančič, "Factors affecting acceptance and use of moodle: an empirical study based on TAM," *Electrical Engineering*, vol. 35, pp. 91–100, 2011.

[44] S. W. Hansen, J. L. Gogan, and R. J. Baxter, "Distributed cognition in geriatric telepsychiatry," in *Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS '12)*, pp. 932–941, Ieee, 2012.

[45] H. S. Park, Y. N. Wu, Y. Ren, and L. Q. Zhang, "A tele-assessment system for evaluating elbow spasticity in patients with neurological impairments," in *Proceedings of the IEEE 10th International Conference on Rehabilitation Robotics (ICORR '07)*, pp. 917–922, Noordwijk, The Netherlands, June 2007.

[46] C. D. Katsis, G. Ganiatsas, and D. I. Fotiadis, "An integrated telemedicine platform for the assessment of affective physiological states," *Diagnostic Pathology*, vol. 1, no. 1, article no. 16, 2006.

[47] F. Koehler, S. Winkler, M. Schieber et al., "Telemedical Interventional Monitoring in Heart Failure (TIM-HF), a randomized, controlled intervention trial investigating the impact of telemedicine on mortality in ambulatory patients with heart failure: study design," *European Journal of Heart Failure*, vol. 12, no. 12, pp. 1354–1362, 2010.

[48] S. Winkler, M. Schieber, S. Lücke et al., "A new telemonitoring system intended for chronic heart failure patients using mobile telephone technology—feasibility study," *International Journal of Cardiology*, vol. 153, no. 1, pp. 55–58, 2011.

[49] S. Capomolla, G. Pinna, M. T. La Rovere et al., "Heart failure case disease management program: a pilot study of home telemonitoring versus usual care," *European Heart Journal, Supplement*, vol. 6, no. 6, pp. F91–F98, 2004.

[50] C. Martínez Alcalá, *Modelo de vistas personalizadas para la gestión de contenido en comunidades I+D+i. [Tesis Doctoral]*, Universidad Politécnica de Cataluña (UPC), 2012.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 598196, 13 pages
http://dx.doi.org/10.1155/2013/598196

Research Article

Mobile Personal Health System for Ambulatory Blood Pressure Monitoring

Luis J. Mena,[1] Vanessa G. Felix,[1] Rodolfo Ostos,[1] Jesus A. Gonzalez,[2] Armando Cervantes,[1] Armando Ochoa,[3] Carlos Ruiz,[3] Roberto Ramos,[1] and Gladys E. Maestre[4,5]

[1] *Department of Computer Engineering, Polytechnic University of Sinaloa, 82199 Mazatlan, SIN, Mexico*
[2] *Department of Computer Science, National Institute of Astrophysics, Optics and Electronics, 72840 Puebla, PUE, Mexico*
[3] *Computer Engineering, Technological Institute of Morelia, 58120 Morelia, MICH, Mexico*
[4] *Institute for Biological Research and Cardiovascular Institute, Faculty of Medicine, University of Zulia, Maracaibo 4002, Venezuela*
[5] *Departments of Psychiatry and Neurology and the Gertrude H. Sergievsky Center, Columbia University, New York, NY 10032, USA*

Correspondence should be addressed to Luis J. Mena; lmena@upsin.edu.mx

Received 10 November 2012; Accepted 12 April 2013

Academic Editor: Angel García-Crespo

The ARVmobile v1.0 is a multiplatform mobile personal health monitor (PHM) application for ambulatory blood pressure (ABP) monitoring that has the potential to aid in the acquisition and analysis of detailed profile of ABP and heart rate (HR), improve the early detection and intervention of hypertension, and detect potential abnormal BP and HR levels for timely medical feedback. The PHM system consisted of ABP sensor to detect BP and HR signals and smartphone as receiver to collect the transmitted digital data and process them to provide immediate personalized information to the user. Android and Blackberry platforms were developed to detect and alert of potential abnormal values, offer friendly graphical user interface for elderly people, and provide feedback to professional healthcare providers via e-mail. ABP data were obtained from twenty-one healthy individuals (>51 years) to test the utility of the PHM application. The ARVmobile v1.0 was able to reliably receive and process the ABP readings from the volunteers. The preliminary results demonstrate that the ARVmobile 1.0 application could be used to perform a detailed profile of ABP and HR in an ordinary daily life environment, bedsides of estimating potential diagnostic thresholds of abnormal BP variability measured as average real variability.

1. Introduction

Chronic noncommunicable diseases (NCDs), such as heart disease, stroke, cancer, chronic respiratory conditions, and diabetes, are the leading cause of mortality in the world, accounting for 63% of all deaths [1]. The leading cause of NCD deaths in 2008 was cardiovascular diseases (CVDs), accounting for 17 million deaths, nearly 30% of global mortality [2]. Thirteen percent of global deaths are attributed to hypertension, the leading risk factor for mortality [2]; the prevalence of hypertension in the global adult population was estimated to be 26% in 2000 and was predicted to increase by about 60% by 2025, to a total of 1.5 billion [3].

Technological innovations that improve prevention and control of CVDs are desperately needed. In this sense, recent studies using the average real variability (ARV) index [4] reported a significant association between high reading-to-reading blood pressure (BP) variability (BPV) and cardiovascular events [4–8]. The ARV is a novel index that has proven to be a more accurate method to assess BPV than the commonly used standard deviation (SD) [4–6, 9, 10]. BPV is a multifaceted phenomenon, influenced by the interaction between external emotional stimuli, such as stress and anxiety, and internal cardiovascular mechanisms that can vary from heartbeat to heartbeat. The complexity of BPV makes analysis difficult, and its independent contribution as

predictor of cardiovascular outcomes is not yet completely clear [11]. Nevertheless, monitoring of BPV in daily life might provide a means to control hypertension, ultimately, preventing CVDs.

Currently available ambulatory blood pressure (ABP) monitors are portable, fully automatic devices that can record BP for 24 hours or longer, while subjects go about their normal daily activities [12]. This technique provides a better estimate of risk in an individual subject than traditional clinical methods, because it (i) removes variability among individual observers, (ii) avoids the "white coat" effect (the transient, but variable elevation of BP in a medical environment [13]) and "masked hypertension" (normotensive by clinical measurement, but hypertensive by ambulatory measurement [14]), (iii) includes the inherent variability of systolic and diastolic BP (SBP and DBP) [15], and (iv) provides information on changes in BP, that is, circadian components. A circadian BP profile with a reduced decrease in nighttime BP level (nondipper status) can indicate increased cardiovascular risk [16–18]. However, the wider use of ABP monitoring, although justified, is limited by its availability and cost. Patients referred to use this technology tend to pay approximately $40 to $70 (USD) per test, depending on the volume of tests performed (i.e., minimum cost per test could only be achieved by a high-volume testing center) [19]. Furthermore, there is currently no experimental or commercial ABP device that estimates BPV through the ARV index.

This paper presents a mobile personal health monitor (PHM) application for ABP monitoring, with the goal of improving health care through early diagnosis of abnormal BP and heart rate (HR) levels, better hypertension control, electronic health registry of individuals, and data for clinical prognosis of CVDs. The paper includes an overview of PHM systems; describes the advantages and limitations of such a mobile health initiative; describes our mobile PHM application in detail, including design considerations and graphic user interface (GUI); discusses the specific context, strengths, and limitations of our approach; proposes future improvements.

2. Mobile Personal Health Monitor

Mobile PHM systems provide personalized, intelligent, reliable, noninvasive, real-time, and pervasive health monitoring [20, 21]. They are part of a body area network (BAN), integrated by a mobile base unit (MBU), with a set of wearable wireless sensors with on-board processing, wireless data transfer, and energy storage capability (Figure 1). The sensors are attached to the user's body; the collected data may be processed locally within the body sensors and/or remotely via wireless transmission to the MBU. The MBU analyzes the data in real-time and provides immediate feedback and personalized information to the user. The analyzed data can also be sent to professional healthcare providers for medical feedback and to support clinical decisions. Mobile PHM systems offer a set of functions that give them global appeal; however, some features might limit their usability and acceptance.

2.1. Pervasive Monitoring. The goal of pervasive monitoring in mobile PHM systems is to provide healthcare services to anyone at any time, overcoming the constraints of place, time, and character [22]. Data processing should be incorporated in the subject's environment in such a way that the interaction between the user and the MBU becomes natural, and the user can obtain personalized information in a totally transparent form.

2.2. Integrated Multisensing Platform. Mobile PHM systems can support biosensors that monitor vital signs (e.g., HR, BP, and blood glucose), environmental conditions (temperature, humidity, and light), and location, integrated by a multisensing platform. Light weight and portability allow long-term, unobtrusive, noninvasive, and ambulatory health monitoring.

2.3. Real-Time Data Analysis. The mobile PHM software of the MBU stores and analyzes the data received from the sensors in real-time, providing instantaneous feedback to the user. PHM systems could alert the user of abnormal events or abrupt changes in near real time, for example, by vibration, loud sound, and/or flashing screen messages.

2.4. Personalized Health Care. Users can configure mobile PHM systems to their specific healthcare needs and preferences, depending on the user's biological profile, as well as the clinical application [23]. Diagnostic thresholds of the risk factor under study could be defined by the user's age, gender, and ethnicity.

2.5. Electronic Data Collection. Mobile PHM systems allow rapid digitization of recorded data, greatly improving quality and efficiency compared to traditional data collection processes with subsequent transcription to computer systems [22]. Continuous monitoring and subsequent transmission to healthcare providers offer an extensive clinical database for data mining analysis of potential risk factors and/or associations among clinical attributes.

2.6. Flexible Communication Protocol. The communication protocols of a mobile PHM system are extremely flexible, because local communication within the BAN can use Wi-Fi [24], Bluetooth [25], or ZigBee [26], and remote communication can be performed via 3G [27] or other available internet communication.

2.7. Data Security, Confidentiality, and Privacy. Concern about security, confidentiality, and privacy of user health data is an important barrier to the use of mobile PHM systems. Security breaches in data transmission could generate situations where confidential data of a heart patient, for example, are manipulated by malicious attackers; regular readings could be altered to indicate a serious problem,

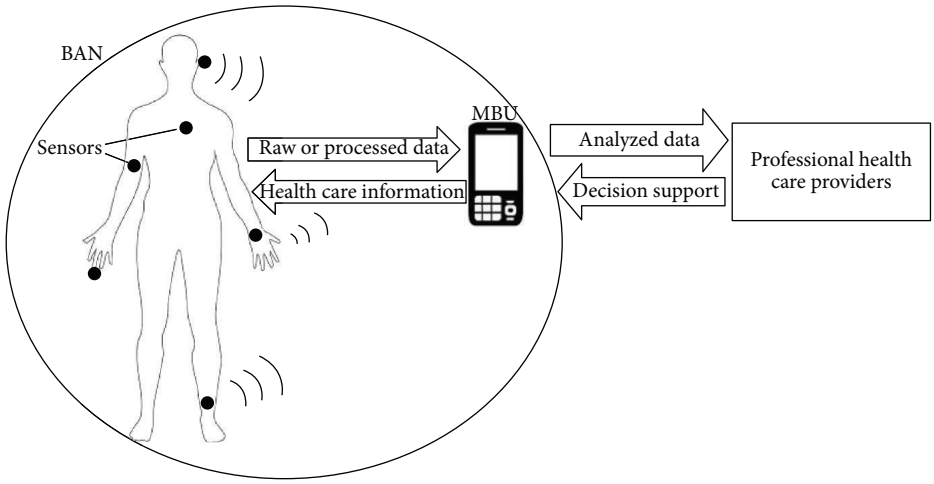


Figure 1: High-level architecture of mobile personal health monitor system.

and the resulting inaccurate feedback could even cause the patient to have a heart attack [22]. Alternatively, the data might be of interest to parties not authorized by the user, such as insurance companies or employers, and such access might result in privacy concerns [28]. Thus, policymakers and program managers must be aware of security issues in the mobile health domain, so that appropriate policies and protections can be developed and implemented [29].

2.8. Need for Multiplatform Applications. The development environments for mobile phones include various systems, such as Android, BlackBerry, iPhone, and Windows Phone platforms. Each uses a different software language, and applications developed for one environment do not operate in the others [30]. For example, Withings company developed a mobile system that manually registers self-measurements of BP through a wired connection between the sensor and the mobile device, but which only operates in Apple platforms [31]. At present, mobile phones service providers do not compete strongly to offer health-related services. However, expectations of market growth [32–34] could worsen the scenario, because the pursuit of market leadership might further fragment the same. Therefore, multiplatform mobile PHM applications are needed.

2.9. Usability and Acceptance among Elderly. While older people are less apt to accept novel and unknown technology than younger people [35], recent studies found that older adults are motivated to use mobile applications if they are satisfactorily informed about the resulting benefits [36, 37]. Elderly people should be the primary target population of mobile PHM for several reasons. First, the global population is aging [38], and chronic NCDs are associated with advanced age [39], so that multimorbidity (coexistence of two or more chronic diseases) is expected to become an increasingly common problem [40]. Second, many elderly people now live alone, with no one to help them to record physical signs like blood pressure [41]. Third, the chances of surviving a fall, heart attack, or stroke are six times greater if the elderly get help within an hour [41]. Finally, mobile PHM systems could provide the elderly with real-time, long-term, nonintrusive assisted living and care services, tailored to their personal health condition [41]. However, acceptance of mobile PHM by older adults is not only based on their health requirements but also on their perspective of technology. Since cognitive performance commonly declines with age, minimizing the complexity of healthcare applications and user-application interactions could be key to the adoption of mobile PHM systems by elderly users and should be considered in their design and development [42]. Therefore, simplicity and motivation seem to be the key factors for usability and acceptance of mobile PHM applications by elderly people.

3. A New Mobile PHM for Ambulatory Blood Pressure Variability

3.1. Rationale. ARV, a recently tested indicator of reading-to-reading BPV [4], has been shown to be significantly related to cardiovascular outcomes [4–8]. ARV attempts to correct for the limitations of the commonly used SD, which accounts only for the dispersion of values around the mean, and not for the order of the BP readings [4–6, 9, 10]. ARV is particularly useful for examining effects of intermittent stress on the cardiovascular system; intermittent BP load on cardiovascular structures may be as important as tonic BP load [43]. Although for most outcomes, ARV was found to be an independent and better predictor of cardiovascular risk than SD [4–6, 8], no previous study has estimated ARV using real-time monitoring, because up to this point, no commercial or experimental real-time device has been capable of monitoring ABP and determining ARV.

Most current healthcare surveillance technologies and diagnostic tools are used in clinical environments, providing only a snapshot of disease under conditions different from the patient's normal lifestyle. The results could lead to difficult or even incorrect diagnoses [20]. Personalized health monitoring in the patient's own environment could

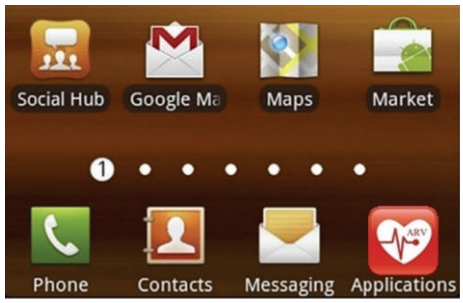


Figure 2: The ARVmobile v1.0. for an Android platform.

significantly improve early detection of CVDs. Personal ABP monitoring would result not only in more accurate diagnosis hypertension but also more effective monitoring of response to therapy, thereby aiding a more tailored approach to patient monitoring.

The ARVmobile v1.0 is a multiplatform, mobile PHM application for real-time, noninvasive, and long-term ABP monitoring (Figure 2). The ARVmobile 1.0 could (i) provide a detailed picture of ABP and heart rate (HR) in a normal environment, (ii) improve the early detection and intervention of hypertension, (iii) improve treatment of hypertension, (iv) identify unexpected responses to antihypertensive treatment, and (v) provide data for the estimation of diagnostic thresholds of abnormal BPV measured as ARV.

3.2. Hardware and Software. The ARVmobile v1.0 was proved in two smartphones (Figure 3). The Samsung I-9100 Galaxy S II, which runs Android OS v2.3 on an ARM Cortex-A9 CPU, with a clock rate of 1.2 GHz Dual Core and 1 GB RAM LPDDR2 [44]. It supports Bluetooth v3.0+HS, Wi-Fi 802.11 a/b/g/n, and USB v2.0; the BlackBerry 9900 Bold that runs BlackBerry OS v7 on a Qualcomm MSM8655 CPU, with a clock rate of 1.2 GHz and 768 MB RAM [45]. It supports Bluetooth 2.1 A2DP/EDR, Wi-Fi 802.11 b/g/n, and USB v2.0. The ARVmobile v1.0 used an ABPM50 [46] to measure BP in millimeters of mercury (mmHg) and HR in beats per minute (bpm). The ABPM50 is an ABP monitor device of low cost that continuously and noninvasively monitors BP level by the oscillometric method [47] and uses Bluetooth for wireless communication. Its sensor has an accuracy of ±3 mmHg and meets ANSI/AAMI SP10-1992 standards [48]. The ABPM50 has a DC power of 3 V (2 AA 1.5V alkali batteries), weighs ~1 kg, and can record more than 600 measurements in 48 hours, in programmable time intervals of 15, 30, 60, 120, or 240 minutes.

The ARVmobile 1.0 for Android was developed in Android 2.2 Froyo [49] with the Eclipse Integrated Development Environment (IDE) 3.7 for Java developers [50] and the Android Development Tool plugin 20.0.2 [51], on MAC OS 10.7 [52], Ubuntu OS 11.10 [53], and Windows OS 7 [54]. The ARVmobile 1.0 for BlackBerry was developed in BlackBerry Java application development 5.0 [55], with the Eclipse IDE 3.6.5 for Java developers [50] and the Blackberry Java plugin 1.5 [56], only on Windows OS 7. Both applications were developed with Java Class Thread [57] as the thread

Figure 3: Android and Blackberry smartphones and the ABPM50 device.

of program execution. The Java Virtual Machine allows an application to have multiple threads of execution running concurrently, so that the smartphone maintains normal operations while receiving real-time BP and HR measurements from the sensor.

3.3. Communication between ABPM50 and Smartphone. Wireless communication between the ABPM50 and smartphone is via Bluetooth. The connection is initiated by setting the smartphone on discovery mode, allowing it to detect the ABPM50 and establish Bluetooth pairing [25]. When the devices detect each other, the smartphone prompts the ABPM50 passkey, sends it to the ABPM50, and validates that the same passkey has been sent and received. If the passkey is verified, secure Bluetooth pairing occurs, and data can be exchanged. To save smartphone energy, the smartphone remembers the Bluetooth pairing and is automatically switched to hidden mode after receiving monitoring readings and changed to discovery mode again before the ABPM50 sends new readings.

The smartphone must know the Media Access Control address of the ABPM50 that uniquely identifies each device of a BAN. Radio frequency communication [25] must be used as transport protocol to share the communication channel with other Bluetooth devices.

3.4. Setting Parameters for the ARVmobile 1.0. Before using the ARVmobile 1.0 to monitor ABP and HR, threshold settings for daytime and nighttime periods must be set correspond to the waking and sleeping patterns of the user (Figure 4). The time intervals between two consecutive readings, which can differ for daytime and nighttime periods, must also be set. According to current clinical guidelines [58, 59], readings should be taken at intervals of 30 minutes or less, and the daytime interval must be smaller than or equal to the nighttime interval. In the current, experimental

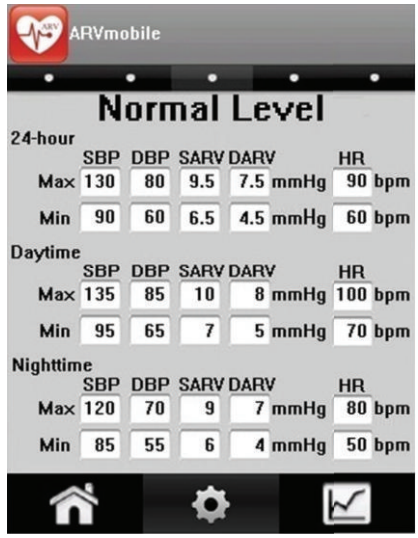


Figure 4: ABP thresholds set on the Blackberry graphic user interface.

version of the ARVmobile, the threshold settings for daytime and nighttime periods of the ARVmobile and ABPM50 must match (Figure 5), so that the smartphone can determine when to enable (receiving readings) and disable (saving energy) the Bluetooth. Although the ARVmobile was designed to use a sensor without on-board processing and to establish a master-slave relationship controlled by the smartphone, current market restrictions did not allow this type of implementation.

3.5. Accuracy of the ARVmobile Monitoring. The goal of the ARVmobile 1.0 and other medical diagnostic devices is to maximize diagnosis accuracy by providing a sufficient amount of accurate data. In ABP monitoring, movement and physical activity often result in invalid ABP readings [60], and recordings can also fail for technical reasons (e.g., improper cuff fitting and auscultatory gap) or patient conditions (e.g., cardiac arrhythmia, rapid pressure changes, severe shock, and HR extremes). A recent review of 25 papers on ABP monitoring found that at least 10% of readings are invalid [61]. Another publication concluded that ABP recordings are successful when at least 85% of readings are suitable for analysis [62]. Current guidelines for management of hypertension indicate that ABP monitoring should be repeated if the recording has fewer than 70% of the expected number of valid values [58]. Thus, there is no specific minimum percentage of valid measurements used to determine the accuracy of ABP monitoring, but there is agreement that a percentage of invalid readings is allowed.

To deal with outlying values, each ABP device has set ranges of BP, ARV, and HR that determine inclusion or exclusion of a reading, which cannot be modified by the user. The ABPM50 has a broader range of operating parameters than other commercial ABPM devices (Table 1), which could increase the number of erroneous readings recorded. Therefore, the measurement thresholds of the ARVmobile 1.0 (Figure 6) were based on those used by the SpaceLabs 90207, which is considered to be an accurate ABP monitor [63–65].

Table 1: Measurement range of different ABP monitors.

Manufacturer	Model	SBP range	DBP range	HR range
CONTEC MS	ABPM50	10–270	10–270	40–240
SunTech	Oscar 2	25–260	25–260	40–200
Well Alynch	6100S	60–250	25–200	40–200
SpaceLabs	90207	60–260	30–200	40–180
A&D Medical	TM-2430-DP	60–280	40–160	30–200

3.6. Estimating ARV with the ARVmobile. The ARVmobile 1.0 computes weighted averages of BP and HR, based on the time intervals between consecutive valid measurements, for 24-hour, daytime, and nighttime periods [6]. Weighted averages provide more accurate estimates than standard averages, which assume that all measurements contribute equally [66]. ARV for a specified period is calculated using

$$\mathrm{ARV} = \frac{1}{\sum w_k} \sum_{k=1}^{n} w_k \times \left|\mathrm{BP}_{k+1} - \mathrm{BP}_k\right|, \quad (1)$$

where n is the number of valid BP readings, k ranges from 1 to n, and w_k is the time interval between BP_k and BP_{k+1}; w_n is the time difference between BP_1 and BP_n.

Using weighted averages and ARV to monitor BPV has several advantages. For example, although the BP recording in Figure 7(a) is visibly more variable than the recording in Figure 7(b), the SD values are the same, because SD reflects only dispersion of values around the mean and not their temporal distribution [67]. In contrast, ARV computed for the more variable recording in Figure 7(a) is almost twice that for the recording in Figure 7(b), providing a more accurate measure of temporal variability in BP.

All of the BP profiles in Figure 7 have a 50% ambulatory SBP load (percentage of systolic readings > 140 mmHg). This ABP parameter improves sensitivity and specificity in the diagnosis of hypertension [68, 69]. Patients with mild hypertension who have an ambulatory SBP load >40% should be strongly considered for antihypertensive therapy [70]. In agreement with this indicator, the weighted average BP (wAvg) for the profile in Figure 7(a) is 136.1 mmHg, which is close to the systolic cutoff for normality in most hypertension guidelines [70], while the standard average BP (Avg) is 125 mmHg, which is between optimal and normal ABP level [71, 72]. Thus, the use of weighted values provides a more accurate hypertension diagnosis.

Figures 7(a) and 7(c) present similar BP profiles, but the record in Figure 7(c) starts with a high SBP and ends with a lower value, similar to a normal circadian, or "dipper," pattern in which BP decreases during sleep and rises sharply upon awakening [73]. Several studies have confirmed that high nocturnal BP, as seen in Figure 7(a), predicts a higher rate of cardiovascular complications [16–18]. This nondipper status is reflected in the higher wAvg in Figure 7(a), because as recommended, the ABP device is programmed with longer nighttime than daytime intervals, resulting in a higher

Blackberry (a) ABPM50 (b)

Figure 5: Matching daytime and nighttime thresholds and recording intervals on a Blackberry and ABPM50 graphic user interface.

Figure 6: Measurement thresholds set on the Android graphic user interface.

weighting factors. The dipper pattern is also observed in Figure 7(d), which has a lower frequency of BP recordings. In this case, wAvg corrects for discarded readings by using higher weighting factors, so that wAvg is similar for Figures 7(c) and 7(d). Although Avg for all four profiles is the same, wAvg computed for Figures 7(c) and 7(d) reflect normal ABP levels and are lower than wAvg for Figure 7(a). Thus, wAvg takes into account effects of BP load, noddipper status, and discarded readings.

3.7. The ARVmobile Feedback. The goals of the ARVmobile 1.0 are to identify abnormal levels of traditional or novel cardiovascular risk factors and provide timely feedback to professional healthcare providers. To accomplish these goals, the ARVmobile 1.0 uses a simple GUI (Figure 8) that is user-friendly for older patients, allowing them access to instantaneous feedback and alerting them about abnormal values that are highlighted in red. The ARVmobile 1.0 also provides a graphic ABP profile (Figure 9) that can be scrolled sideways on the screen. The user has the option of sending ABP results via e-mail to specified third parties, such as family members, and sends a complete output report in PDF format to specified healthcare professionals (Figure 10). The report includes 24-hour, daytime, and nighttime BP and HR levels, ARV, BP load, pulse pressure (differences between SBP and DBP [74]), and identification of abnormal APB values outside of the set thresholds. A graphic presentation of the circadian pattern (Figure 9) can be integrated into the PDF file or sent separately in JPGE format. The ARVmobile 1.0 allows feedback to be forwarded to up to three e-mail addresses and can be configured to automatically send ABP and HR profiles at the end of each monitoring period or only after monitoring periods in which abnormal values are detected. The report can be sent in Spanish or English (Figure 11).

3.8. The ARVmobile Trial. At the time of writing this paper, over twenty-one subjects aged higher than 51 (mean age 58.9 ± 6.1 years; 61.9% women) have used the ARVmobile 1.0. Volunteer subjects without history of cardiovascular disease and hypertension were recruited in the mobile computer laboratory of the Polytechnic University of Sinaloa, in Mazatlan, Mexico. Recruited participants had reasonable computer skills and were assertive about using new technologies. Informed consent was obtained from every participant. Participants were provided of an ABPM50 device and an Android or Blackberry smartphones and were instructed

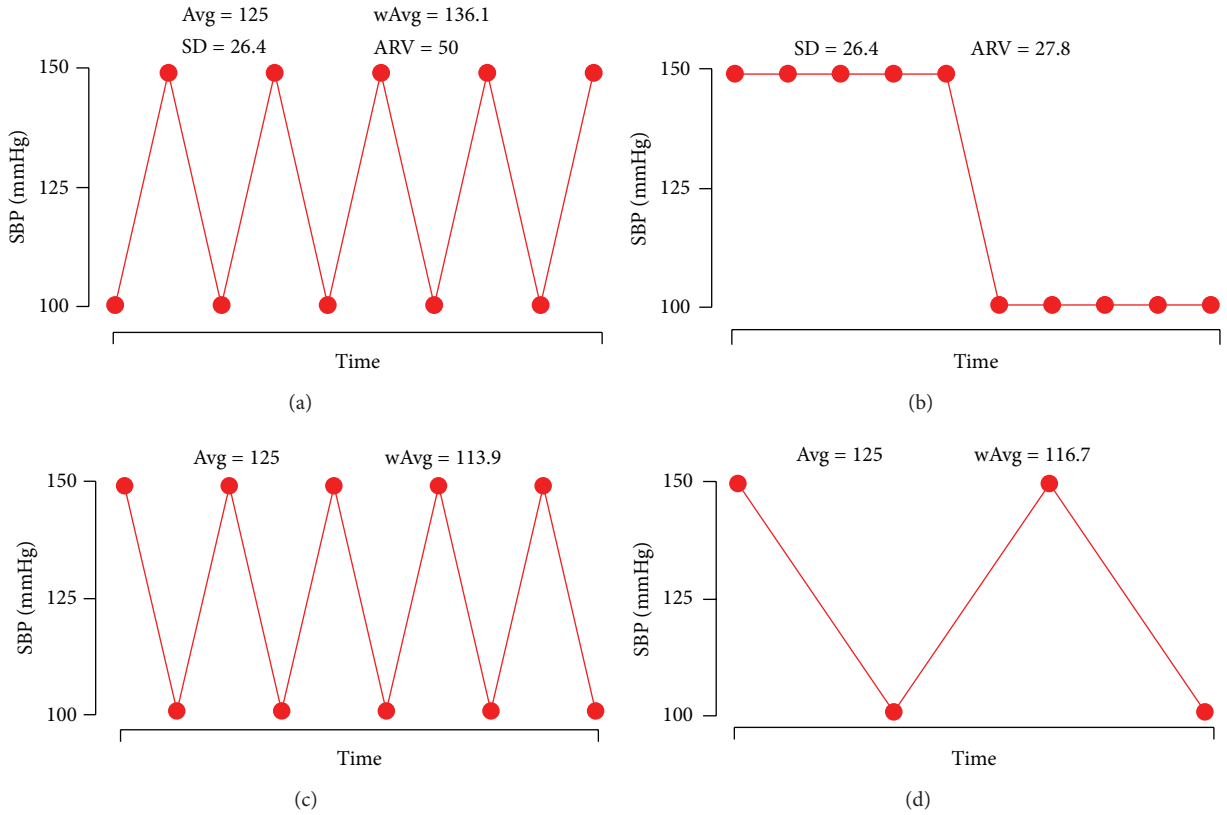


Figure 7: Estimates of average BP (Avg), weighted average BP (wAvg), standard deviation (SD), and average real variability (ARV) for four different systolic blood pressure (SBP) profiles, recorded at intervals of 15 min. The number of values is lower in (d), due to invalid readings that were not recorded. In all four cases, Avg = 125 mmHg.

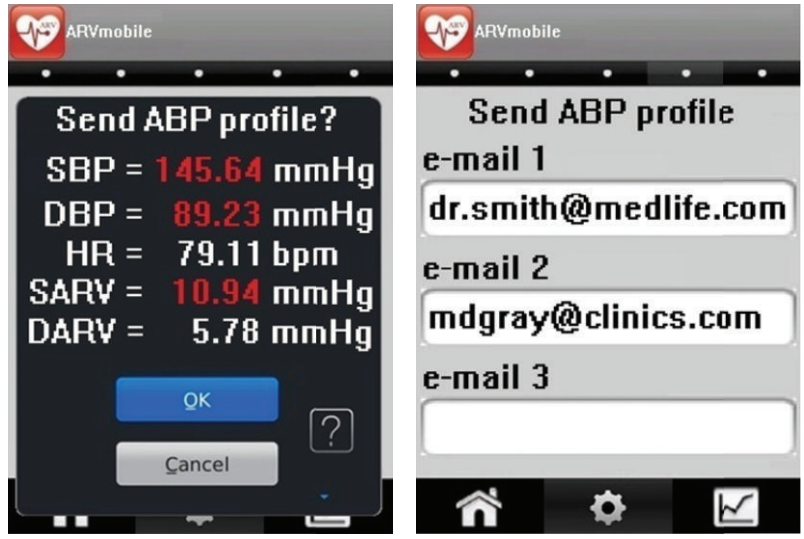
how to use the PHM application and ABP sensor and to continue with their usual daily activities. The ABPM50 was programmed to obtain readings at 15 minute intervals during awake time (06:00–22:59) and at 30 minutes intervals for the sleeping period (23:00–05:59). The median number of ABP readings was 70 (5th to 95th percentile, 62–79) with a high percentage (>75%) of valid readings. To validate the accuracy of the wireless communication and data processing, BP and HR readings from each subject recorded in the PHM system were compared against those captured by the ABP monitor, without finding inconsistencies between the ABPM50 and ARVmobile records. A posterior survey indicated that the majority of the participants found the ARVmobile 1.0 easy to use and considered the time spent learning how to use the PHM application reasonable. The preliminary results demonstrate that the ARVmobile 1.0 application could be used to perform a detailed profile of ABP and HR in an ordinary daily life environment.

4. Discussion

The ARVmobile 1.0 is the result of an interdisciplinary effort in clinical research, data mining analysis, and development of mobile applications. Development focused on (i) an innovative mobile PHM application; (ii) support of early diagnosis and intervention of CVDs, based on cardiovascular risk factors and pattern recognition [75–78] identified by clinical studies [4, 79–82]; (iii) novel strategies to improve adoption by user-friendliness for elderly people.

The inclusion of the capability of calculation of ARV is a major innovation of the ARVmobile 1.0. Although ARV has been shown to be an accurate indicator of BPV [4–6, 9, 10], as well as a significant and independent predictor of cardiovascular complications, when adjusted for BP level and other covariates [4–8], its use to date has been primarily for research, rather than for clinical purposes. However, the clinical use of BPV via calculation of ARV could improve the diagnosis and prognosis of hypertension by providing better information on progressive and end-point organ damage associated with high BP values. Furthermore, BPV could be useful in assessing the efficacy of antihypertensive agents [83] and developing new therapeutic drugs to treat hypertension [84].

The use of weighted averages (wAvg) to construct BP and HR profiles is another major innovation of the ARVmobile 1.0, and wAvg takes the sequential order of measurements into account and, therefore, more accurately represent effects of intermittent stress on the cardiovascular system. Effects of intermittent BP load on cardiovascular structures might be as important as tonic BP load [43], and estimation of ARV and wAvg incorporates this phenomenon.



(a) (b)

Figure 8: The ARVmobile-Blackberry graphic user interface for (a) instantaneous feedback and (b) forwarding an ABP profile to selected recipients.

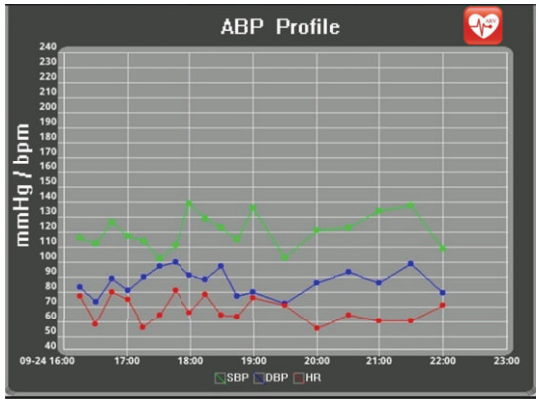


Figure 9: The ARVmobile-Blackberry graphic user interface showing a circadian pattern in ABP.

An important goal of the ARVmobile 1.0 is to provide more effective personal surveillance of elderly hypertensive patients, resulting in a personally tailored approach to hypertension control. One previous study showed that people who check their own BP tend to be more conscious of the importance of taking their medication on a regular basis [85]. The user might notice resistance to antihypertensive drugs or specific foods and activities that raise their BP, and they can adjust their lifestyle, accordingly.

Although use of mobile PHMs could improve and support the lives of the elderly, seniors are often viewed as luddites, reluctant to use modern technology. Thus, development of the ARVmobile 1.0 considered user acceptance and adoption specifically with respect to elderly populations. The first barrier is hardware reluctance. However, a recent market study reported that the 6.2 million working Americans aged 65 and older increased smartphone ownership by 150% during 2007–2010 [86]. The acceptability and ethical aspects of body sensors also could be an important consideration to elderly users. However, widespread use of other personal technologies, such as the Bluetooth headset and smartphone, suggests that older people will easily adjust to wearing sensors for long periods outside of clinical environments [20]. The ARVmobile 1.0 was purposefully designed to be user-friendly to elderly patients with reduced vision and manual dexterity. It has a simplified GUI with a bright screen, large text and numbers, and simple input buttons with touchscreen technology, all of which have been proven to be efficient for older adults [87].

The usability of any application depends on the acquisition of new procedural knowledge for proper operation and interaction; therefore, simple design makes the application more usable [88]. This is especially true for older users, because cognitive performance commonly slows down with age. The ARVmobile 1.0 has only three interactive menus (home, settings, and graphical trend). Only the settings menu contains submenus, and these are usually adjusted only once. Security mechanisms, such as user identification for access, were omitted, because users can activate such mechanisms through the security settings of the smartphone. To assure privacy, reports forwarded to selected recipients lack personal identification, which is already associated with the source e-mail address. The ability to operate the ARVmobile 1.0 in either Spanish or English is another user-friendly feature (Figure 11).

Perhaps the most important factor in acceptance and adoption of a mobile PHM system is the user's motivation, which depends on their understanding of the magnitude of the health problem and the benefits of the mobile application. To motivate potential users of the ARVmobile 1.0, it is necessary to disseminate information about the importance of prevention and control of hypertension. Although global

ARVmobile v.1.0

Date: 09/24/2012 Duration: 23:42 min Readings: 73 (91.25%)
Daytime: 07:00 Nighttime: 23:00 End: 06:42
Daytime interval: 15 min Nighttime interval: 30 min

Ambulatory blood pressure monitoring

Normal level
24-hour: SBP: 130–90 mmHg; DBP: 80–60 mmHg; SARV: 9.5–6.5 mmHg; DARVD: 7.5–4.5 mmHg; HR: 90–60 bpm
Daytime: SBP: 135–95 mmHg; DBP: 85–65 mmHg; SARV: 10–7 mmHg; DARVD: 7–4 mmHg; HR: 100–70 bpm
Nighttime: SBP: 120–85 mmHg; DBP: 70–55 mmHg; SARV: 9–6 mmHg; DARVD: 7–4 mmHg; HR: 80–50 bpm

	Unit	24 hours			Daytime			Nighttime		
		SBP	DBP	HR	SBP	DBP	HR	SBP	DBP	HR
Readings		73	73	73	59	59	59	14	14	14
Weighted average	(mmHg/bpm)	130.27	*84.95	87.64	*136.91	*89	96.83	*121.12	*79.36	74.97
ARV	(mmHg)	*12.37	*8.79	—	12.12	7.91	—	*12.71	*10.02	—
Pulse pressure	(mmHg)	46.37		—	47.59		—	41.21		—
Blood pressure load	(%)	56.16	79.45	—	54.24	74.58	—	35.71	78.57	—
Maximum value	(mmHg/bpm)	162	110	139	162	110	139	132	88	87
Minimum value	(mmHg/bpm)	96	56	37	114	67	37	96	56	64

* Ambulatory blood pressure value at least 1 unit outside of normal level.
— Indicates not applicable.
• Systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (HR) levels were computed with weighted average, using as weighting factor the time interval between consecutive valid readings.
• Average real variability (ARV) was computed with weighted average of the absolute differences of consecutive valid readings.
• Pulse pressure is the standard average of the differences between systolic and diastolic pressures recorded at the same time.
• Blood pressure load is the percentage of blood pressure readings above than normal level.

Figure 10: Report forwarded by the ARVmobile to specified healthcare providers.

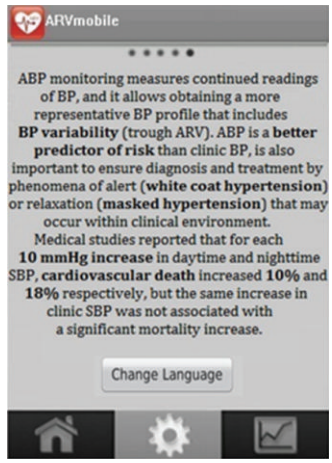


Figure 11: ARVmobile-Android graphic user interface allowing change of language.

prevalence of hypertension is high [2], more than 50% of hypertensive individuals are unaware of their condition [89], and control of hypertension is only around 13.5% [90]. Awareness of isolated systolic hypertension (ISH), which involves high SBP and normal DBP and is more common in the elderly [91], is particularly appropriate to the ARVmobile 1.0. ISH is a major cardiovascular risk factor: each 20/10 mmHg increase doubles the risk for older hypertensive patients [92]. In older patients with ISH, ABP is a better predictor of risk than clinical BP measurements [93, 94]. Cardiovascular death increased 10% and 18%, respectively, for each 10 mm Hg increase in daytime and nighttime SBP, but the same increase in clinical SBP was not associated with a significant mortality increase [93]. The ARVmobile 1.0 presents some of this information in the language change option (Figure 11).

Use of the ARVmobile 1.0 could help avoid overtreatment of elderly patients with white coat hypertension or lack of treatment for those with masked hypertension [95]. Both issues are associated with cost effectiveness, as well as diagnosis and treatment. The cost of hypertension control based on conventional clinical BP measurements can be up to four times higher than control based on ABP monitoring [96], due to unnecessary drug therapy for patients with white coat hypertension [97, 98]; in addition the cost-benefit ratio would be expected to increase, as the cost of managing hypertension rises with increasing rates of diagnosis and prescribing of new, more expensive antihypertensive [62]. Furthermore, the actual raising trend to increase the development of low-cost medical devices [99], and the constant rise of physician fees for primary care [100], could also promote the use of personal ABP monitoring, because in clinical practice, traditional ABP monitors need to be attached and detached by skilled medical technologists [101].

The main limitation of usability and adoption of the ARVmobile 1.0 is the ABP sensor design, which must integrate wearability, accurate measurement, power source miniaturization, low power use in reading biosignals and wireless transmission, and secure data transference. To deliver truly personalized health care, the biosensors must be invisible to the user, avoiding activity restriction or behavior modification [20]. Biosensors should be small and lightweight, which depends largely on the size and weight of batteries. However, battery capacity is directly proportional to size [102]. The gold standard for measuring BP with ABP devices is the oscillometric method, which requires inflating a cuff around the arm and requires high power. Development of cuffless

sensors that use other BP measurement techniques, such as pulse arrival time [103, 104], but still provide accuracy, is necessary. Another option would be to integrate the sensor into nonclothing items that are already worn by patients. The cost of such novel technology could emerge as the new limitation.

Future development of the ARVmobile requires defining the minimum number of ABP readings required to assess BPV in a reliable or reproducible manner, even with conventional ABP monitors such number is not known [82]. Several studies found that short-term, reading-to-reading BPV estimated by ABP monitoring had poor reproducibility compared to ABP level [105–108], which could account for the rather diverse findings regarding the clinical value of BPV as a predictor of cardiovascular outcomes. The reproducibility of ARV could be improved by increasing the number of BP readings, but 24-hour ABP monitoring can be uncomfortable, especially for elderly patients. Therefore, it is necessary to determine a minimum range of BP readings to calculate ARV without significant loss of information, because accurate calculation of ARV with an adequate number of BP measurements might have great relevance for clinical purposes, and its final implementation in the ARVmobile 1.0 would improve the adoption and use of PHM applications for CVDs prevention. In this sense, the ARVmobile could be a useful and cost-saving PHM system to perform a detailed profile of ABP and HR in an ordinary daily life environment and to estimate potential diagnostic thresholds of abnormal BPV measured as ARV.

Conflict of Interests

The authors do not have a direct financial relation with the commercial identity mentioned in their paper that might lead to a conflict of interests for any of them.

Acknowledgments

This paper was supported by the Secretaria de Educación Pública, Mexico DF, México (PROMEP/103-5/11/6951 and PROMEP/103-5/11/4145). The Maracaibo Aging Study was funded by the Venezuelan Grant FONACIT G-97000726, Fundaconciencia and by award no. R01AG036469 from the National Institute on Aging. The authors acknowledge the support of the Programa Interinstitucional para el Fortalecimiento de la Investigación y el Posgrado del Pacífico, Programa Delfín, which integrates 69 universities (public and private), technological institutes (federal and state), and state councils of science and technology in several states of Mexico. The authors are grateful to the Instituto Tecnológico de Morelia for its institutional cooperation and collaboration. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes or any other funding.

References

[1] A. Alwan, D. R. MacLean, L. M. Riley et al., "Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries," *The Lancet*, vol. 376, no. 9755, pp. 1861–1868, 2010.

[2] World Health Organization, *Global Health Risks—Mortality and Burden of Disease Attributable to Selected Major Risk*, World Health Organization, Geneva, Switzerland, 2009.

[3] P. M. Kearney, M. Whelton, K. Reynolds, P. Muntner, P. K. Whelton, and J. He, "Global burden of hypertension: analysis of worldwide data," *The Lancet*, vol. 365, no. 9455, pp. 217–223, 2005.

[4] L. Mena, S. Pintos, N. V. Queipo, J. A. Aizpúrua, G. Maestre, and T. Sulbarán, "A reliable index for the prognostic significance of blood pressure variability," *Journal of Hypertension*, vol. 23, no. 3, pp. 505–511, 2005.

[5] S. D. Pierdomenico, M. Di Nicola, A. L. Esposito et al., "Prognostic value of different indices of blood pressure variability in hypertensive patients," *American Journal of Hypertension*, vol. 22, no. 8, pp. 842–847, 2009.

[6] T. W. Hansen, L. Thijs, Y. Li et al., "Prognostic value of reading-to-reading blood pressure variability over 24 hours in 8938 subjects from 11 populations," *Hypertension*, vol. 55, no. 4, pp. 1049–1057, 2010.

[7] P. Veerabhadrappa, K. M. Diaz, D. L. Feairheller et al., "Enhanced blood pressure variability in a high cardiovascular risk group of African Americans: FIT4Life Study," *Journal of the American Society of Hypertension*, vol. 4, no. 4, pp. 187–195, 2010.

[8] G. Schillaci, G. Bilo, G. Pucci et al., "Relationship between short-term blood pressure variability and large-artery stiffness in human hypertension: findings from 2 large databases," *Hypertension*, vol. 60, no. 2, pp. 369–377, 2012.

[9] G. Parati and D. Rizzoni, "Assessing the prognostic relevance of blood pressure variability: discrepant information from different indices," *Journal of Hypertension*, vol. 23, no. 3, pp. 483–486, 2005.

[10] K. Eguchi, S. Hoshide, Y. Hoshide, S. Ishikawa, K. Shimada, and K. Kario, "Reproducibility of ambulatory blood pressure in treated and untreated hypertensive patients," *Journal of Hypertension*, vol. 28, no. 5, pp. 918–924, 2010.

[11] T. W. Hansen, Y. Li, and J. A. Staessen, "Blood pressure variability remains an elusive predictor of cardiovascular outcome," *American Journal of Hypertension*, vol. 22, no. 1, pp. 3–4, 2009.

[12] T. G. Pickering, D. Shimbo, and D. Hass, "Ambulatory blood pressure monitoring," *The New England Journal of Medicine*, vol. 354, pp. 2368–2374, 2006.

[13] T. G. Pickering, G. D. James, C. Boddie, G. A. Harshfield, S. Blank, and J. H. Laragh, "How common is white coat hypertension?" *Journal of the American Medical Association*, vol. 259, no. 2, pp. 225–228, 1988.

[14] T. Pickering, K. Davidson, W. Gerin, and J. E. Schwarts, "Masked hypertension," *Hypertension*, vol. 40, pp. 795–796, 2002.

[15] A. Frattola, G. Parati, C. Cuspidi, F. Albini, and G. Mancia, "Prognostic value of 24-hour blood pressure variability," *Journal of Hypertension*, vol. 11, no. 10, pp. 1133–1137, 1993.

[16] T. W. Hansen, J. Jeppesen, S. Rasmussen, H. Ibsen, and C. Torp-Pedersen, "Ambulatory blood pressure and mortality: a population-based study," *Hypertension*, vol. 45, no. 4, pp. 499–504, 2005.

[17] E. Ingelsson, K. Björklund, L. Lind L, J. Ärnlöv, and J. Sundström, "Diurnal blood pressure pattern and risk of congestive heart failure," *The Journal of the American Medical Association*, vol. 295, no. 24, pp. 2859–2866, 2006.

[18] G. Mancia, R. Facchetti, M. Bombelli, G. Grassi, and R. Sega, "Long-term risk of mortality associated with selective and combined elevation in office, home, and ambulatory blood pressure," *Hypertension*, vol. 47, no. 5, pp. 846–853, 2006.

[19] Ontario Health Technology Advisory Committee, *OHTAC Recommendation: Twenty-Four-Hour Ambulatory Blood Pressure Monitoring in Hypertension*, Ontario Health Technology Advisory Committee, Ontario, Canada, 2012.

[20] O. Aziz, B. Lo, J. Pansiot, L. Atallah, G. Z. Yang, and A. Darzi, "From computers to ubiquitous computing by 2010: health care," *Philosophical Transactions of the Royal Society A*, vol. 366, no. 1881, pp. 3805–3811, 2008.

[21] P. Leijdekkers, V. Gay, and E. Barin, "Feasibility study of a non invasive cardiac rhythm management system," *International Journal of Assistive Robotics and Systems*, vol. 10, no. 4, pp. 5–14, 2009.

[22] R. Shahriyar, F. Bari, G. Kundu, S. I. Ahamed, and M. Akbar, "Intelligent mobile health monitoring system (IMHMS)," *International Journal of Control and Automation*, vol. 2, no. 3, pp. 13–27, 2009.

[23] V. Gay, P. Leijdekkers, and E. Barin, "A Mobile rehabilitation application for the remote monitoring of cardiac patients after a heart attack or a coronary bypass surgery," in *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '09)*, June 2009.

[24] Wi-Fi Alliance, *Securing Wi-Fi Wireless Networks with Today's Technologies*, Wi-Fi Alliance, Austin, Tex, USA, 2003.

[25] P. Bhagwat, "Bluetooth: technology for short-range wireless apps," *IEEE Internet Computing*, vol. 5, no. 3, pp. 96–103, 1995.

[26] P. Baronti, P. Pillai, V. W. C. Chook, S. Chessa, A. Gotta, and Y. F. Hu, "Wireless sensor networks: a survey on the state of the art and the 802.15.4 and ZigBee standards," *Computer Communications*, vol. 30, no. 7, pp. 1655–1695, 2007.

[27] G. Camarillo and M. Garcia-Martin, *The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds*, John Wiley & Sons, Chichester, UK, 3rd edition, 2008.

[28] J. Barnickel, H. Karahan, and U. Meyer, "Security and privacy for mobile electronic health monitoring and recording systems," in *Proceedings of 11th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pp. 1–6, June 2010.

[29] World Health Organization, *MHealth New Horizons for Health Through Mobile Technologies*, World Health Organization, Geneva, Switzerland, 2011.

[30] K. Patrick, W. G. Griswold, F. Raab, and S. S. Intille, "Health and the mobile phone," *American Journal of Preventive Medicine*, vol. 35, no. 2, pp. 177–181, 2008.

[31] M. Essany, "mHealth: iPhones to provide mobile means for monitoring blood pressure," Mobile Marketing Watch, 2011, http://www.mobilemarketingwatch.com/mhealth-iphones-to-provide-mobile-means-for-monitoring-blood-pressure-16425/.

[32] PricewaterhouseCoopers, *Emerging mHealth: Paths for Growth*, PricewaterhouseCoopers, London, UK, 2012.

[33] MobiHealthNews, *State of the Industry: Mobile Health Q1 2012*, MobiHealthNews, 2012.

[34] Research2Guidance, *Mobile Health Market Report 2011–2016—The Impact of Smartphone Applications on the Mobile Health Industry*, Research2Guidance, London, UK, 2012.

[35] C. Hertzog and M. K. Bleckley, "Age differences in the structure of intelligence influences of information processing speed," *Intelligence*, vol. 29, no. 3, pp. 191–217, 2001.

[36] E. D. Mynatt, A. S. Melenhorst, A. D. Fisk, and W. A. Rogers, "Aware technologies for aging in place: understanding user needs and attitudes," *IEEE Pervasive Computing*, vol. 3, no. 2, pp. 36–41, 2004.

[37] A. S. Melenhorst, W. A. Rogers, and D. G. Bouwhuis, "Older adults' motivated choice for technological innovation: evidence for benefit-driven selectivity," *Psychology and Aging*, vol. 21, no. 1, pp. 190–195, 2006.

[38] The Centers for Disease Control and Prevention, "Public health and aging: trends in aging—United States and worldwide," *The Journal of the American Medical Association*, vol. 289, no. 11, pp. 1371–1373, 2003.

[39] E. M. Crimmins, "Trends in the health of the elderly," *Annual Review of Public Health*, vol. 25, pp. 79–98, 2004.

[40] A. Marengoni, B. Winblad, A. Karp, and L. Fratiglioni, "Prevalence of chronic diseases and multimorbidity among the elderly population in Sweden," *American Journal of Public Health*, vol. 98, no. 7, pp. 1198–1200, 2007.

[41] Z. Lv, F. Xia, G. Wu, L. Yao, and Z. Chen, "iCare: a mobile health monitoring system for the elderly," in *Proceedings of the 2010 IEEE International Conference on Green Computing and Communications*, pp. 699–705, 2010.

[42] M. Ziefle and S. Bay, "How older adults meet complexity: aging effects on the usability of different mobile phones," *Behaviour and Information Technology*, vol. 24, no. 5, pp. 375–389, 2005.

[43] M. F. O'Rourke, "Basic concepts for the understanding of large arteries in hypertension," *Journal of Cardiovascular Pharmacology*, vol. 7, pp. S14–S21, 1985.

[44] Samsung Galaxy S II, http://www.samsung.com/global/microsite/galaxys2/html/specification.html.

[45] BlackBerry Bold 9900, http://us.BlackBerry.com/smartphones/BlackBerry-bold-9900-9930.html#/h:/smartphones//BlackBerry-bold-9900-9930/phone-specifications.html.

[46] Contec Medical System ABPM50, http://www.contecmed.com/main/product_show.asp?ArticleID=549.

[47] CAS Medical Systems, *Performance of the CAS Oscillometric Algorithm When Compared Against Various Commercially Available NIBP Simulators*, CAS Medical Systems, Branford, Conn, USA, 2007.

[48] Association for the Advancement of Medical Instrumentation, 1992, http://www.techstreet.com/cgi-bin/detail?doc_no=aami%7Csp 10_92,product_id=15999.

[49] Platforms/Android Developers, http://developer.android.com/tools/revisions/platforms.html.

[50] Mylyn Download Archives, http://www.eclipse.org/mylyn/downloads/archive.php.

[51] ADT Plugin/Android Developers, http://developer.android.com/tools/sdk/eclipse-adt.html.

[52] Apple Support Lion, http://www.apple.com/support/lion/.

[53] Ubuntu, http://www.ubuntu.com/download/desktop.

[54] Windows 7, http://windows.microsoft.com/is-IS/windows7/get-know-windows-7.

[55] BlackBerry JDE 5. 0. 0 API Reference, http://www.BlackBerry.com/developers/docs/5.0.0api/index.html.

[56] Blackberry, *BlackBerry Java Plug-in for Eclipse Version: 1.5*, Blackberry, Redwood City, Calif, USA, 2006.

[57] Thread (Java 2 Platform SE v1. 4. 2), http://docs.oracle.com/javase/1.4.2/docs/api/java/lang/Thread.html.

[58] G. Mancia, G. De Backer, A. Dominiczak et al., "2007 guidelines for the management of arterial hypertension," *European Heart Journal*, vol. 28, no. 12, pp. 1462–1536, 2007.

[59] *What Is Hypotension?*, National Heart Lung Blood Institute, 2010, http://www.nhlbi.nih.gov/health/health-topics/topics/hyp/.

[60] B. P. McGrath, "Ambulatory blood pressure monitoring position statement," *The Medical Journal of Australia*, vol. 176, no. 12, pp. 588–592, 2002.

[61] T. Rechciński, "A survey of recent reports on ambulatory blood pressure monitoring," *World Journal of Hypertension*, vol. 2, no. 1, pp. 7–12, 2012.

[62] G. A. Head, B. P. McGrath, A. S. Mihailidou et al., "Ambulatory blood pressure monitoring in Australia: 2011 consensus position statement," *Journal of Hypertension*, vol. 30, no. 2, pp. 253–266, 2012.

[63] E. O'Brien, F. Mee, N. Atkins, and K. O'Malley, "Accuracy of the SpaceLabs 90207 determined by the British Hypertension Society protocol," *Journal of Hypertension*, vol. 9, supplement 5, pp. S25–S31, 1991.

[64] P. Baumgart and J. Kamp, "Accuracy of the SpaceLabs Medical 90217 ambulatory blood pressure monitor," *Blood Pressure Monitoring*, vol. 3, no. 5, pp. 303–307, 1998.

[65] Spacelabs Medical Inc, (Nasdaq:SLMD) today reported financial results for the second quarter ended July 1, 2000, Health/Medical Writers, 2007, http://www.thefreelibrary.com/Spacelabs+Medical, Inc.+Reports+Second+Quarter+Results.-a063724759.

[66] T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions," *Technometrics*, vol. 37, no. 2, pp. 285–294, 1995.

[67] R. W. de Boer, J. M. Karemaker, and J. Strackee, "On the spectral analysis of blood pressure variability," *The American Journal of Physiology*, vol. 21, no. 3, pp. H685–H687, 1986.

[68] R. C. Hermida and D. E. Ayala, "Evaluation of the blood pressure load in the diagnosis of hypertension in pregnancy," *Hypertension*, vol. 38, no. 3, pp. 723–729, 2001.

[69] P. K. Zachariah, S. G. Sheps, D. M. Ilstrup et al., "Blood pressure load—a better determinant of hypertension," *Mayo Clinic Proceedings*, vol. 63, no. 11, pp. 1085–1091, 1988.

[70] W. B. White, H. M. Dey, and P. Schulman, "Assessment of the daily blood pressure load as a determinant of cardiac function in patients with mild-to moderate hypertension," *American Heart Journal*, vol. 118, no. 4, pp. 782–795, 1989.

[71] T. W. Hansen, M. Kikuya, L. Thijs et al., "Diagnostic thresholds for ambulatory blood pressure moving lower: a review based on a meta-analysis-clinical implications," *Journal of Clinical Hypertension*, vol. 10, no. 5, pp. 377–381, 2008.

[72] M. Kikuya, T. W. Hansen, L. Thijs et al., "Diagnostic thresholds for ambulatory blood pressure monitoring based on 10-year cardiovascular risk," *Circulation*, vol. 115, no. 16, pp. 2145–2152, 2007.

[73] K. Shimada, K. Kario, Y. Umeda, S. Hoshide, Y. Hoshide, and K. Eguchi, "Early morning surge in blood pressure," *Blood Pressure Monitoring*, vol. 6, no. 6, pp. 349–353, 2001.

[74] A. W. Haider, M. G. Larson, S. S. Franklin, and D. Levy, "Systolic blood pressure, diastolic blood pressure, and pulse pressure as predictors of risk for congestive heart failure in the Framingham Heart study," *Annals of Internal Medicine*, vol. 138, no. 1, pp. 10–16, 2003.

[75] L. Mena, J. A. Gonzalez, and G. Maestre, "Extracting new patterns for cardiovascular disease prognosis," *Expert Systems*, vol. 26, no. 5, pp. 364–377, 2009.

[76] L. J. Mena, E. E. Orozco, V. G. Felix, R. Ostos, J. Melgarejo, and G. E. Maestre, "Machine learning approach to extract diagnostic and prognostic thresholds: application in prognosis of cardiovascular mortality," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 750151, 6 pages, 2012.

[77] L. Mena and J. A. Gonzalez, "Symbolic one-class learning from imbalanced datasets: application in medical diagnosis," *International Journal on Artificial Intelligence Tools*, vol. 18, no. 2, pp. 273–309, 2009.

[78] L. Mena and J. A. Gonzalez, "Machine learning for imbalanced datasets: application in medical diagnostic," in *Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference (FLAIRS '06)*, pp. 574–579, May 2006.

[79] G. E. Maestre, G. Pino-Ramírez, A. E. Molero et al., "The Maracaibo Aging study: population and methodological issues," *Neuroepidemiology*, vol. 21, no. 4, pp. 194–201, 2002.

[80] G. Maestre, L. Mena, G. Pino-Ramirez et al., "Incidence of dementia in the Maracaibo Aging study," *Alzheimer's & Dementia*, vol. 7, supplement 4, pp. S356–S357, 2011.

[81] L. Mena, J. D. Melgarejo, C. Chavez et al., "Relevance of blood pressure variability among the elderly: findings from the Maracaibo Aging study," *Journal of Hypertension*, vol. 29, p. e312, 2011.

[82] K. Stolarz-Skrzypek, L. Thijs, T. Richart et al., "Blood pressure variability in relation to outcome in the international database of ambulatory blood pressure in relation to cardiovascular outcome," *Hypertension Research*, vol. 33, no. 8, pp. 757–766, 2010.

[83] Y. Zhang, D. Agnoletti, M. E. Safar, and J. Blacher, "Effect of antihypertensive agents on blood pressure variability: the natrilix SR versus candesartan and amlodipine in the reduction of systolic blood pressure in hypertensive patients (X-CELLENT) study," *Hypertension*, vol. 58, no. 2, pp. 155–160, 2011.

[84] D. F. Su and C. Y. Miao, "Reduction of blood pressure variability: a new strategy for the treatment of hypertension," *Trends in Pharmacological Sciences*, vol. 26, no. 8, pp. 388–390, 2005.

[85] M. A. Weber, *Hypertension Medicine*, Humana Press, New Jersey, NJ, USA, 2000.

[86] "6. 2 million seniors are working past retirement—and buying smartphones and HDTVs," Aging Online, 2011, http://www.aging-online.com/62-million-seniors-are-working-past-retiremen.

[87] A. Holzinger, "User-centered interface design for disabled and elderly people: first experiences with designing a patient communication system (PACOSY)," in *Computers Helping People With Special Needs*, vol. 2398 of *Lecture Notes in Computer Science*, pp. 34–41, Springer, Berlin, Germany, 2002.

[88] A. Holzinger, G. Searle, and A. Nischelwitzer, "On some aspects of improving mobile applications for the elderly," in *Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction (UAHCI '07)*, pp. 923–932, 2007.

[89] A. Chockalingam, "World hypertension day and global awareness," *The Canadian Journal of Cardiology*, vol. 24, no. 6, pp. 441–444, 2008.

[90] M. Pereira, N. Lunet, A. Azevedo, and H. Barros, "Differences in prevalence, awareness, treatment and control of hypertension between developing and developed countries," *Journal of Hypertension*, vol. 27, no. 5, pp. 963–975, 2009.

[91] A. V. Chobanian, "Isolated systolic hypertension in the elderly," *The New England Journal of Medicine*, vol. 357, no. 8, pp. 789–796, 2007.

[92] J. R. Chiong, W. S. Aronow, I. A. Khan et al., "Secondary hypertension: current diagnosis and treatment," *International Journal of Cardiology*, vol. 124, no. 1, pp. 6–21, 2008.

[93] J. A. Staessen, L. Thijs, R. Fagard et al., "Predicting cardiovascular risk using conventional versus ambulatory blood pressure in older patients with systolic hypertension," *Journal of the American Medical Association*, vol. 282, no. 6, pp. 539–546, 1999.

[94] L. M. Wing, M. A. Brown, L. J. Beilin, P. Ryan, and C. M. Reid, "Reverse white-coat hypertension'in older hypertensive," *Journal of Hypertension*, vol. 20, no. 4, pp. 639–644, 2002.

[95] G. Rodriguez-Roca, F. J. Alonso-Moreno, A. Garcia-Jimenez et al., "Cost-effectiveness of ambulatory blood pressure monitoring in the follow-up of hypertension," *Blood Pressure*, vol. 15, no. 1, pp. 27–36, 2006.

[96] B. Ewald and B. Pekarsky, "Cost analysis of ambulatory blood pressure monitoring in initiating antihypertensive drug treatment in Australian general practice," *Medical Journal of Australia*, vol. 176, no. 12, pp. 580–583, 2002.

[97] L. Aitken and C. Addison, "The cost-effectiveness of ambulatory blood pressure monitoring," *Professional nurse*, vol. 12, no. 3, pp. 198–202, 1996.

[98] M. L. Burr, E. Dolan, E. W. O'Brien, E. T. O'Brien, and P. McCormack, "The value of ambulatory blood pressure in older adults: the Dublin outcome study," *Age and Ageing*, vol. 37, no. 2, pp. 201–206, 2008.

[99] Trade and Labor and Israel Export & International Cooperation Institute, *Market Analysis Report: China's Medical Device and Healthcare IT Industries, Israel Ministry of Industry*, Trade and Labor and Israel Export & International Cooperation Institute, Tel Aviv, Israel, 2010.

[100] S. Zuckerman and D. Goin, *How Much Will Medicaid Physician Fees for Primary Care Rise in 2013? Evidence from a 2012 Survey of Medicaid Physician Fees*, Kaiser Commission on Medicaid and the Uninsured, Menlo Park, Calif, USA.

[101] H. Nakano, M. Kikuya, A. Hara et al., "Self-monitoring of ambulatory blood pressure by the Microlife WatchBP O3—an application test," *Clinical and Experimental Hypertension*, vol. 33, no. 1, pp. 34–40, 2011.

[102] A. Milenković, C. Otto, and E. Jovanov, "Wireless sensor networks for personal health monitoring: issues and an implementation," *Computer Communications*, vol. 29, no. 13-14, pp. 2521–2533, 2006.

[103] C. C. Y. Poon and Y. T. Zhang, "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society*, pp. 5877–5880, September 2005.

[104] F. S. Cattivelli and H. Garudadri, "Noninvasive cuffless estimation of blood pressure from pulse arrival time and heart rate with adaptive calibration," in *Proceedings of the 6th International Workshop on Wearable and Implantable Body Sensor Networks (BSN '09)*, pp. 114–119, Berkeley, Calif, USA, June 2009.

[105] L. Thijs, J. Staessen, R. Fagard, P. Zachariah, and A. Amery, "Number of measurements required for the analysis of diurnal blood pressure profile," *Journal of Human Hypertension*, vol. 8, no. 4, pp. 239–244, 1994.

[106] P. Palatini, P. Mormino, C. Canali et al., "Factors affecting ambulatory blood pressure reproducibility: results of the HARVEST trial," *Hypertension*, vol. 23, no. 2, pp. 211–216, 1994.

[107] B. Wizner, D. G. Dechering, L. Thijs et al., "Short-term and long-term repeatability of the morning blood pressure in older patients with isolated systolic hypertension," *Journal of Hypertension*, vol. 26, no. 7, pp. 1328–1335, 2008.

[108] M. Di Rienzo, G. Grassi, A. Pedotti, and G. Mancia, "Continuous versus intermittent blood pressure measurements in estimating 24-hour average blood pressure," *Hypertension*, vol. 5, no. 2, pp. 264–269, 1983.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 651091, 6 pages
http://dx.doi.org/10.1155/2013/651091

Research Article

Statistical Evaluation of a Fully Automated Mammographic Breast Density Algorithm

Mohamed Abdolell,[1,2,3] Kaitlyn Tsuruda,[2] Gerry Schaller,[1,2] and Judy Caines[1,2,4]

[1] *Department of Diagnostic Radiology, Dalhousie University, Halifax, NS, Canada B3H 2Y9*
[2] *Department of Diagnostic Imaging, Capital District Health Authority, Halifax, NS, Canada B3H 2Y9*
[3] *Division of Medical Education/Informatics, Dalhousie University, Halifax, NS, Canada B3H 2Y9*
[4] *Nova Scotia Breast Screening Program, Halifax, NS, Canada B3H 2Y9*

Correspondence should be addressed to Mohamed Abdolell; mo@dal.ca

Received 2 October 2012; Revised 5 April 2013; Accepted 9 April 2013

Academic Editor: Giner Alor-Hernández

Visual assessments of mammographic breast density by radiologists are used in clinical practice; however, these assessments have shown weaker associations with breast cancer risk than area-based, quantitative methods. The purpose of this study is to present a statistical evaluation of a fully automated, area-based mammographic density measurement algorithm. Five radiologists estimated density in 5% increments for 138 "For Presentation" single MLO views; the median of the radiologists' estimates was used as the reference standard. Agreement amongst radiologists was excellent, ICC = 0.884, 95% CI (0.854, 0.910). Similarly, the agreement between the algorithm and the reference standard was excellent, ICC = 0.862, falling within the 95% CI of the radiologists' estimates. The Bland-Altman plot showed that the reference standard was slightly positively biased (+1.86%) compared to the algorithm-generated densities. A scatter plot showed that the algorithm moderately overestimated low densities and underestimated high densities. A box plot showed that 95% of the algorithm-generated assessments fell within one BI-RADS category of the reference standard. This study demonstrates the effective use of several statistical techniques that collectively produce a comprehensive evaluation of the algorithm and its potential to provide mammographic density measures that can be used to inform clinical practice.

1. Introduction

Breast density refers to fibroglandular tissue in the breast and is one of the top major risk factors for breast cancer. Women with extremely dense breasts (75% or greater mammographic density) have a four- to sixfold increase in the risk of developing breast cancer compared to those with fatty breasts (less than 25% density) [1–3].

Traditionally, visual assessment by radiologists has been used to characterize and quantify mammographic density (and a woman's risk for breast cancer) using Wolfe Grades, Tabar Patterns, Boyd Scales, or the American College of Radiologists' (ACR) Breast Imaging Reporting and Data System (BI-RADS) density lexicon [4–7]. Despite good reproducibility, methods used to characterize mammographic density have shown weaker associations with breast cancer risk compared to methods quantifying mammographic density [2, 3, 8] and suffer from inter- and intraobserver variability.

The ACR has stated that radiologists' visual assessments of percent breast density using the BI-RADS lexicon are "not reliably reproducible" [9]. This fundamental lack of reproducibility has led to the development of various semi- and fully automated algorithms to quantify percent breast density as a means to overcome inter- and intraobserver variability. It is therefore important to apply rigorous statistical methods to evaluate the performance of these algorithms.

1.1. State of the Art. Area-based methods used to quantify mammographic density have produced reliable and standardized mammographic density measurements on a continuous scale. The de facto standard of such methods is the Cumulus software [10, 11]. Using Cumulus, a digitized film-screen mammogram is displayed and a trained user selects a threshold value to separate the breast area from the background (i.e., the region of interest). A second threshold is

then selected to identify regions of dense breast tissue, and the percent breast density is calculated as the area of dense tissue divided by the area of the region of interest. Despite being a proven predictor of breast cancer risk, the semiautomated nature of Cumulus' breast density assessments is susceptible to inter- and intraobserver variability and could be improved by a fully automated method. Additionally, this software is intended for use with digitized film-screen mammograms. As 90% of certified mammography units in the USA are now full-field digital [12], a software for use with full-field digital mammograms (FFDMs) is needed.

Volume-based methods theoretically yield accurate estimates of mammographic density and so it is simply assumed that volume-based density estimates are associated with breast cancer risk, as has been demonstrated to be the case for area-based estimates (both visually and algorithmically assessed) [13, 14]. Volumetric methods use "For Processing" FFDMs and DICOM header information to calculate density. Yet, volume-based estimates have not been shown to demonstrate a similarly strong association with breast cancer risk [11, 15, 16]. Additionally, the underlying distribution of mammographic density estimates from volumetric methods is significantly more left-skewed than that of area-based methods (typical range 0–40% versus 0–100%) [17], making them difficult to interpret by radiologists, who are not simply able to visualize mammographic density as a volumetric construct [11, 15].

The assessment of the agreement between percent breast density algorithms and an expert radiologist should necessarily quantify the consistency or reproducibility of measurements made by these two "raters" on the same set of digital mammograms. The intraclass correlation coefficient (ICC) provides such a measure of agreement [18]. The Bland-Altman plot is another way to assess agreement between raters. Scatter and box plots can also yield insights into the level of agreement between raters. Yet, much of the literature validating emerging density measurement algorithms relies on the use of the Pearson correlation coefficient, ρ, which is a measure of the linear dependence between two raters and can be quite high despite the agreement being poor [18, 19]. Overall percent agreement is another statistic that is used to assess agreement but is also flawed as it does not factor in any inherent inter- and intrarater variability [19]. Reporting of a single numerical measure of agreement alone is one-dimensional and does not present a comprehensive perspective on algorithm performance.

This paper presents several statistical methods that collectively provide a more comprehensive evaluation of the performance of a fully automated area-based image analysis algorithm that generates percent breast density measures from FFDMs.

2. Materials and Methods

138 "For Presentation" FFDMs collected from the Capital District Health Authority in Nova Scotia were retrospectively analyzed. Images were acquired on Siemens full-field digital mammography machines and automatically postprocessed by the manufacture's proprietary software at the time of acquisition. This early stage work has focused on the mediolateral oblique views and excluded craniocaudal views as it has been shown in the literature that mammographic density estimates from only one view are sufficient to indicate breast cancer risk [20]. In addition, the ACRs' National Mammography Database breast density element definition stipulates that "if left and right breasts differ, use the higher density" [21].

2.1. Percent Density Analysis. Percent mammographic density was measured by a fully automated research-based algorithm that uses "For Presentation" FFDMs to calculate an area-based measure of density as a percentage on a continuous scale (Figure 1, Panels 1(a) through 1(d)). Using view position and image laterality information from the DICOM header (elements (0018, 5101) and (0020, 0062), resp.) the software creates and applies a mask to identify the breast envelope (region of interest) by removing the pectoral muscle, subcutaneous fat, and overlay text (Panel 1(b)). A variation of the MaxEntropy and Moments thresholding methods is applied to determine a threshold for dense tissue in the breast [22, 23]. The area of the dense tissue (i.e., the number of pixels of dense tissue) is then calculated (Panel 1(c)), as is the area of the region of interest (i.e., the number of pixels in the breast area, Panel 1(d)), and the final density estimate is calculated as the ratio of dense tissue area to the region of interest. In this manner, the software uniquely generates a reproducible, fully automated, area-based estimate of mammographic density using "For Presentation" FFDM images.

To evaluate the agreement between the algorithm and an expert mammographer, percent mammographic density was visually assessed by five radiologists in 5% density increments (0%, 5%, . . . , 95%, 100%) using five megapixel Barco Screens supported by the Syngo MammoReport Software (VB24D, Siemens AS). Visual assessments were performed by two senior mammographers, one junior mammographer, one senior resident, and one fellow. This 21-point scale was used as a proxy for a continuous measure.

2.2. Statistical Analysis. To quantify the reliability of estimates performed by the radiologists' visual assessments, Intraclass Correlation Coefficients (ICCs) were used to measure interobserver agreement. Although the interpretation of ICCs can vary depending on the context, the ICC is equivalent to a quadratically weighted Kappa, and a widely referenced scale for interpretation of Kappa values can be used as a general guide [24, 25]. Specifically, ICC values of 0.00–0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.00 were used to indicate poor, fair, moderate, substantial, and excellent to perfect agreement, respectively.

It has repeatedly been shown that radiologists' visual assessments of mammographic density are associated with breast cancer risk [1, 3, 4, 10, 26]. As such, the median of the visual assessments performed by the five participating radiologists was considered to be the reference standard for this analysis. The algorithm was considered promising in informing clinical practice if the agreement between the

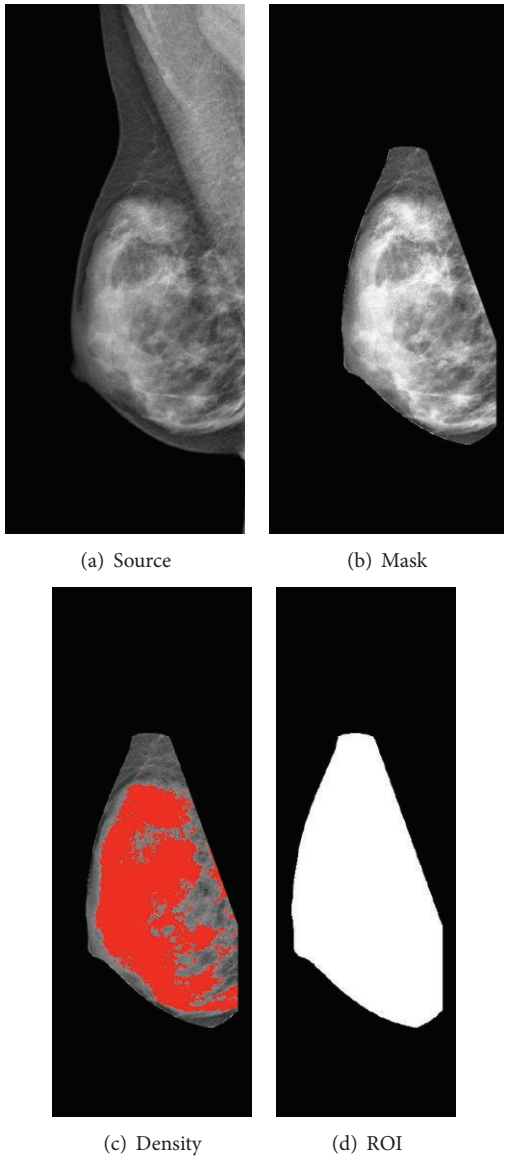


Figure 1: A sequence of processed images generated at various steps of the algorithm for estimating area-based mammographic density: (a) a "For Presentation" mammogram from our sample; (b) the image after a mask has been applied to identify the breast envelope; (c) the area of dense tissue (red pixels); and (d) the region of interest as a binary map of the breast envelope. The algorithm calculates percent breast density as the number of red pixels in Panel (c) divided by the number of white pixels in Panel (d).

algorithm and the reference standard fell within the 95% CI of the ICC of the radiologists.

The ICC was used to quantify the level of agreement between the algorithm and the reference standard, and a scatterplot was used to demonstrate the relationship between the two. A Bland-Altman difference plot was used to analyze the agreement between the algorithm and the reference standard and to quantify the amount and direction of bias as well as the upper and lower limits of agreement (bias $\pm 1.96\sigma$ of the difference) [27]. Lastly, a box-and-whisker plot was used to visualize the results in terms of the BI-RADS density lexicon (0–24%, 25–49%, 50–74%, and 75–100%) [7].

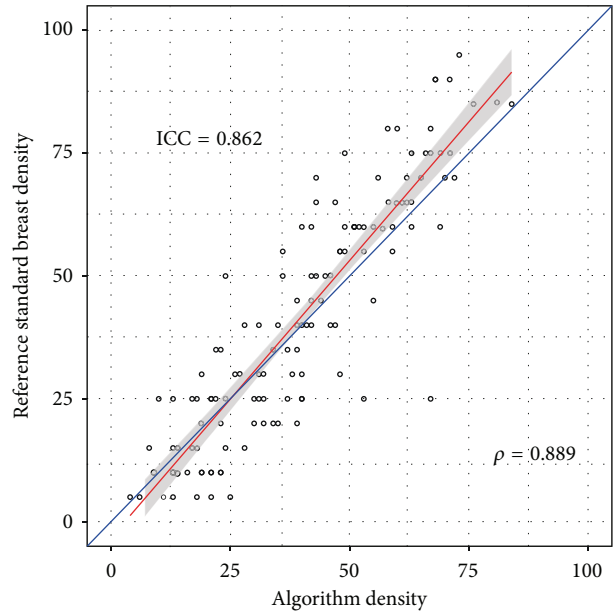


Figure 2: Scatter plot showing the relationship between the mammographic density estimates produced by the algorithm (x-axis) and the reference standard (y-axis). The blue line indicates perfect agreement between the algorithm and the reference standard, in which case all points would fall exactly on the line of agreement. The red line is the line of the best fit determined by linear least squares regression analysis and shows that the algorithm tends to slightly overestimate density compared to the reference standard for lower densities and slightly underestimate density compared to the reference standard for higher densities.

3. Results

Five radiologists visually assessed 138 images to estimate mammographic density, and the algorithm was applied to those same 138 images to generate a fully automated density assessment for each of the images.

The radiologists' visual assessments were in excellent agreement with an ICC = 0.884, 95% CI (0.854, 0.910). The algorithm demonstrated excellent agreement with the reference standard with an ICC = 0.862, which fell within the 95% CI of the agreement between the radiologists' visual assessments. The algorithm is validated well on an external set of 261 mammograms, ICC = 0.841.

The Pearson correlation coefficient between the algorithm and the reference standard assessments was $\rho = 0.889$.

The algorithm slightly overestimated low densities and underestimated high densities compared to the reference standard (Figure 2). Overall, there was a small, positive bias in the reference standard assessments compared to the algorithm assessments, as measured by the mean difference between the reference standard and the algorithm assessments (bias = 1.86%) (Figure 3). Additionally, the upper and lower agreement levels were both less than 25%, and thus

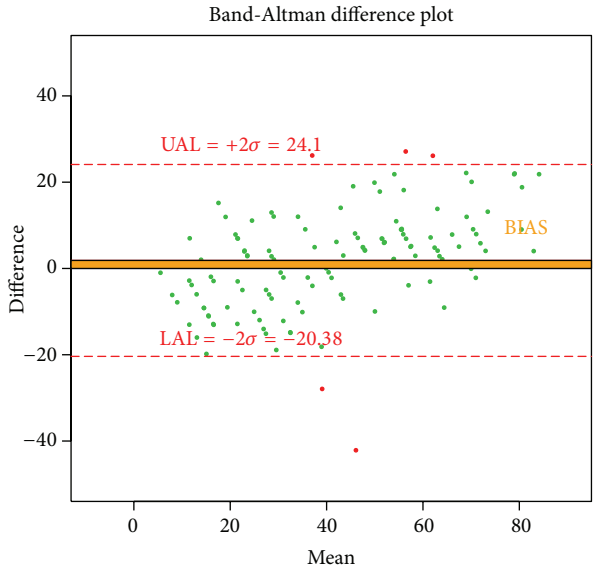


Figure 3: Bland-Altman difference plot showing agreement between the algorithm and the reference standard measures of mammographic density. The difference refers to the reference standard minus the algorithm assessment. The absolute values of the upper and lower agreement limits are <25%, which is the span of a single category in the 4-level BI-RADS density classification scheme. A bias of +1.86%, as indicated by the orange band above the horizontal zero difference line, shows that the reference standard density is on average only slightly higher than the density generated by the algorithm.

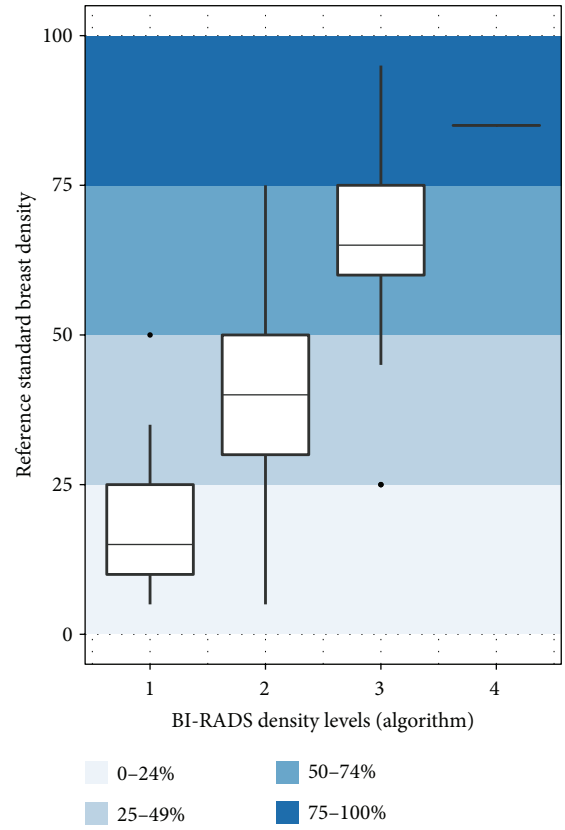


Figure 4: Box-and-whisker plot displaying the distribution of reference standard mammographic density assessments falling into the algorithm-derived classifications designated by the standard 4-level BI-RADS density lexicon. Ideally, each of the boxes and their whiskers should be entirely contained in their respective BI-RADS levels. The reference standard mammographic density assessments in the lowest and the highest BI-RADS levels are well classified, while the middle two levels overlap in both directions into adjacent BI-RADS levels.

approximately 95% of the data classified by the algorithm was within one BI-RADS category of the reference standard classification (Figure 3).

When the algorithm and reference standard estimates were classified using the BI-RADS density lexicon, the box-and-whisker plots showed good agreement within categories (Figure 4). Each box was contained in the accordant colour bar, and, as expected from the Bland-Altman difference plot, the tails on the graphs did not exceed the adjacent BI-RADS categories.

4. Discussion

The algorithm demonstrates excellent agreement with radiologists' visual assessments of mammographic density. Critically, the observed magnitude of this agreement falls within the 95% CI of agreement observed between radiologists. This algorithm is unique in that it generates fully automated mammographic density measurements that can be straightforwardly compared with visually determined radiologists' estimates, which are well accepted as being associated with breast cancer risk.

The sole use of the Pearson correlation coefficient (ρ) provides a one-dimensional and overinflated impression of the level of agreement.

The statistical evaluation presented in this paper used ICCs and Bland-Altman, scatter, and box plots to quantify agreement and bias in breast density assessment between a fully automated algorithm and radiologists' assessments. This multifaceted methodology can be employed to comprehensively evaluate the performance of any breast density measurement algorithm and provides an alternative to the often reported Pearson correlation coefficient and percent agreement statistics which do not consider random chance agreement and cannot quantify bias between different raters.

As breast density legislation gains momentum in the USA and mammography providers are required to disclose breast density in the lay report, there will be an increasing need for automated solutions that provide reliable and accurate measurements of breast density. A woman's breast density will be used to determine her optimal followup, and thus the performance of these algorithms must be evaluated using robust statistical methodologies.

5. Conclusion

Further work is needed to extend the applicability of the breast density algorithm to FFDMs from other manufacturers

as each manufacturer has their own proprietary image processing algorithms that generate "For Presentation" images. Additionally, as radiologists use both mediolateral and craniocaudal views to assess breast density in a clinical setting, the present algorithm must also be extended to accommodate the analysis of craniocaudal views.

The present algorithm is an effective research tool and shows promise in its ability to provide automated mammographic density measurements that can be used to inform clinical practice. The Pearson correlation coefficient (ρ) provides an inadequate, inflated, and overoptimistic measure of the level of agreement. The statistical methods employed provide a comprehensive evaluation of the level of agreement between the algorithm and the reference standard and confirm that the algorithm has an excellent level of agreement with the reference standard. Agreement between raters can only be adequately assessed using multiple statistical methods.

Ethical Approval

This study was approved by the Capital District Health Authority Research Ethics Board: CDHA-RS/2007-365.

Conflict of Interests

Mohamed Abdolell is the founder of Densitas Inc. and is a shareholder in the company. Kaitlyn Tsuruda is an employee at the company.

Acknowledgments

The authors would like to thank Dr. Melanie McQuaid, Dr. Ieva Klavina, and Dr. Chris Lightfoot for providing visually assessed mammographic density measurements used in this study. This research was supported by the Department of Diagnostic Radiology, Dalhousie University and Capital District Health Authority, and the Canadian Breast Cancer Foundation (Atlantic Region). The study sponsors had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the paper; nor in the decision to submit the paper for publication.

References

[1] N. F. Boyd, H. Guo, L. J. Martin et al., "Mammographic density and the risk and detection of breast cancer," *New England Journal of Medicine*, vol. 356, no. 3, pp. 227–236, 2007.

[2] J. Brisson, C. Diorio, and B. Mâsse, "Wolfe's parenchymal pattern and percentage of the breast with mammographic densities: redundant or complementary classifications?" *Cancer Epidemiology Biomarkers and Prevention*, vol. 12, no. 8, pp. 728–732, 2003.

[3] V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiology Biomarkers and Prevention*, vol. 15, no. 6, pp. 1159–1169, 2006.

[4] J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," *Cancer*, vol. 37, no. 5, pp. 2486–2492, 1976.

[5] I. T. Gram, E. Funkhouser, and L. Tabár, "The Tabar classification of mammographic parenchymal patterns," *European Journal of Radiology*, vol. 24, no. 2, pp. 131–136, 1997.

[6] N. F. Boyd, H. M. Jensen, G. Cooke, and H. L. Han, "Relationship between mammographic and histological risk factors for breast cancer," *Journal of the National Cancer Institute*, vol. 84, no. 15, pp. 1170–1179, 1992.

[7] *Breast Imaging Reporting and Data Systems (BIRADS)*, American College of Radiology, Reston, Va, USA, 1993.

[8] M. Garrido-Estepa, F. Ruiz-Perales, J. Miranda et al., "Evaluation of mammographic density patterns: reproducibility and concordance among scales," *BMC cancer*, vol. 10, supplement 485, 2010.

[9] American College of Radiology, "ACR Statement on Reporting Breast Density in Mammography Reports and Patient Summaries," 2012, http://www.acr.org/About-Us/Media-Center/Position-Statements.

[10] N. F. Boyd, J. W. Byng, R. A. Jong et al., "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *Journal of the National Cancer Institute*, vol. 87, no. 9, pp. 670–675, 1995.

[11] N. Boyd, L. Martin, A. Gunasekara et al., "Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes," *Cancer Epidemiology Biomarkers and Prevention*, vol. 18, no. 6, pp. 1754–1762, 2009.

[12] FDA, "MQSA National Statistics," 2013, http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandards-ActandProgram/FacilityScorecard/ucm113858.htm.

[13] O. Alonzo-Proulx, N. Packard, J. M. Boone et al., "Validation of a method for measuring the volumetric breast density from digital mammograms," *Physics in Medicine and Biology*, vol. 55, no. 11, pp. 3027–3044, 2010.

[14] M. J. Yaffe, "Mammographic density. Measurement of mammographic density," *Breast Cancer Research*, vol. 10, no. 3, article 209, 2008.

[15] D. Kontos, P. R. Bakic, R. J. Acciavatti, E. F. Conant, and A. D. A. Maidment, "A comparative study of volumetric and area-based breast density estimation in digital mammography: results from a screening population," in *Digital Mammography*, vol. 6136 of *Lecture Notes in Computer Science*, pp. 378–385, 2010.

[16] J. Ding, R. Warren, I. Warsi et al., "Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: case-control study," *Cancer Epidemiology Biomarkers and Prevention*, vol. 17, no. 5, pp. 1074–1081, 2008.

[17] R. Highnam, S. Brady, M. Yaffe, N. Karssemeijer, and J. Harvey, "Robust breast composition measurement-volpara TM," in *Proceedings of the 10th International Conference on Digital Mammography (IWDM '10)*, vol. 6136 of *Lecture Notes in Computer Science*, pp. 342–349, 2010.

[18] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 1, no. 8476, pp. 307–310, 1986.

[19] R. J. Hunt, "Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability," *Journal of Dental Research*, vol. 65, no. 2, pp. 128–130, 1986.

[20] J. Stone, J. Ding, R. M. L. Warren, and S. W. Duffy, "Predicting breast cancer risk using mammographic density measurements from both mammogram sides and views," *Breast Cancer Research and Treatment*, vol. 124, no. 2, pp. 551–554, 2010.

[21] American College of Radiology, "National mammography database data elements," version 2.0, 2009, https://nrdr.acr.org/Portal/HELP/NMD/nmd_data_elements.pdf.

[22] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, & Image Processing*, vol. 29, no. 3, pp. 273–285, 1985.

[23] W. H. Tsai, "Moment-preserving thresholding: a new approach," *Computer Vision, Graphics, & Image Processing*, vol. 29, no. 3, pp. 377–393, 1985.

[24] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educational and Psychological Measurement*, vol. 33, pp. 613–619, 1973.

[25] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[26] J. A. Harvey and V. E. Bovbjerg, "Quantitative assessment of mammographic breast density: relationship with breast cancer risk," *Radiology*, vol. 230, no. 1, pp. 29–41, 2004.

[27] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 1, no. 8476, pp. 307–310, 1986.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 796387, 17 pages
http://dx.doi.org/10.1155/2013/796387

Research Article

Development of an Expert System as a Diagnostic Support of Cervical Cancer in Atypical Glandular Cells, Based on Fuzzy Logics and Image Interpretation

Karem R. Domínguez Hernández,[1] Alberto A. Aguilar Lasserre,[1] Rubén Posada Gómez,[1] José A. Palet Guzmán,[2] and Blanca E. González Sánchez[1]

[1] *Instituto Tecnológico de Orizaba, Avenida Oriente 9 No. 852, Colonia Emiliano Zapata, 94320 Orizaba, VER, Mexico*
[2] *Hospital Regional de Río Blanco, Entronque Autopista Orizaba-Puebla km 2, 94735 Río Blanco, VER, Mexico*

Correspondence should be addressed to Karem R. Domínguez Hernández; karem_dguez@hotmail.com

Received 26 October 2012; Revised 16 January 2013; Accepted 17 February 2013

Academic Editor: Alejandro Rodríguez González

Cervical cancer is the second largest cause of death among women worldwide. Nowadays, this disease is preventable and curable at low cost and low risk when an accurate diagnosis is done in due time, since it is the neoplasm with the highest prevention potential. This work describes the development of an expert system able to provide a diagnosis to cervical neoplasia (CN) precursor injuries through the integration of fuzzy logics and image interpretation techniques. The key contribution of this research focuses on atypical cases, specifically on atypical glandular cells (AGC). The expert system consists of 3 phases: (1) risk diagnosis which consists of the interpretation of a patient's clinical background and the risks for contracting CN according to specialists; (2) cytology images detection which consists of image interpretation (IM) and the Bethesda system for cytology interpretation, and (3) determination of cancer precursor injuries which consists of in retrieving the information from the prior phases and integrating the expert system by means of a fuzzy logics (FL) model. During the validation stage of the system, 21 already diagnosed cases were tested with a positive correlation in which 100% effectiveness was obtained. The main contribution of this work relies on the reduction of false positives and false negatives by providing a more accurate diagnosis for CN.

1. Introduction

During the last fifty years, the incidence and mortality rate for cervical cancer has decreased, mostly in developed countries thanks to the implementation of prevention programs (cytology). However, among gynecologic types of cancer, this pathology ranks second place in developing countries, mainly, in poorer zones.

On the other hand, the risk factors linked to this disease are of the socioeconomic type, primarily consisting of multiple incidences in rural zones, as well as the young age of the patients. The human papillomavirus is considered to be a causal agent and is linked to the early beginning of sexual relations, especially with casual sex partners.

In 2008, more than 80,000 women were diagnosed with cervix cancer and 36,000 died because of this disease in Latin America only. In developing countries, such as Mexico, CN is linked to an even higher mortality rate, contrary to rates in more developed countries. However, in Mexico, several studies have been performed in this field. Every year, the Mexican Institute of Social Security (IMSS in Spanish) diagnoses around 15,000 cases of CN [1].

This current research is based on specialists' expertise in diagnosing cervix cancer. The expert system models the information provided by doctors for the decision-making process in the diagnosis. Then, in a second phase, the system performs an analysis by segmenting the processed images to show the interest parameters in the cytological image suggested by the Bethesda system [2] for its diagnosis. In the first two phases, the system is integrated by a fuzzy logics model and its graphic interface in which a cytological image is uploaded and the interest data about the clinical

records of a patient is typed out. Finally, a diagnosis is suggested for precursor injuries of CN. In cases dealing with doctor's expertise, an expert system that is able to diagnose positive or negative cases does not rise any interest; for this reason, the contribution of this work focuses on the atypical, where the information is not enough to emit either a positive or a negative CN diagnosis. These kinds of cases are known as atypical glandular cells (AGC). The expert system described herein is able to offer a diagnosis since it has mathematical information from the image processing, which a doctor cannot obtain when it is analyzed manually. With this information and the addition of the clinical records of a patient gathered in an expert system, it can be used as a support tool in the decision-making process for the specialists. As a result, a diagnosis for precursor injuries for atypical cases is obtained.

This work is structured in three parts. The first stage describes the doctor's perspective, where the studied illness is presented, as well as its factors and the diagnosis' methods. Furthermore, in the same stage, a review of the related literature in this topic is presented along with its main contributions. The second stage presents the development of this research among the three stages that comprise the expert system. Finally, in the third stage, a description of the case study is presented, and the tests are run on the already diagnosed cases in order to validate the system.

2. Medic Perspective and a Literature Review

After breast cancer, worldwide, cervical cancer is the second most frequent cancer among women. Until 2008, an estimated of 529,800 deaths were related to this disease around the world [3]. Nowadays, cytological screening by means of Pap smear has reduced the cancer incidences in terminal stages. As a result, an important increase in precursor cancer injuries in asymptomatic women has been recorded. In their natural evolution, these injuries can be treated and eradicated if an appropriate diagnosis is available on time.

Everyday, medical breakthroughs provide valuable information that boosts treatments and cures to a countless number of illnesses. However, the need for creation of multidisciplinary groups, which can provide different knowledge and perspectives about the same topic, is also evident.

2.1. Cancer Manifestation. Cancer is a pathologic tissue growth originated due to a continual proliferation of abnormal cells. It can stem from any type of cell from different tissues in the human organism.

There are many types of cancer. One of the most threatening to women is cervical cancer (CN), which unlike the others, if it is detected during its early stages, the possibilities for a total recovery are plenty [4]. More than 80% of deaths caused by CN, come from countries with high or medium levels of poverty. It is to be expected that the annual figure of deaths caused by CN will raise and exceed to 11 million by 2030 [5].

Cervical cancer's underlying cause is the infection by the human papillomavirus (HPV), which is a common sexually transmitted disease (STD). However, 10 or 20 years are needed for a precursor injury produced by HPV to be turned into invasive cancer. Unfortunately, it is estimated that 95% of women (in the reproductive age) who inhabit developing countries have never taken a Pap smear [6].

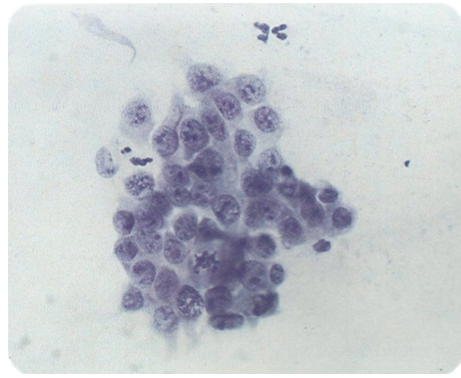


Figure 1: AGC.

The total recovery rate for this disease is closely associated to the stage of development at the moment of the diagnosis and to the availability of its treatment, since CN is deadly if it remains untreated. Due to its complexity, CN treatment directly relies on its appropriate and accurate diagnosis.

There is no question about the decisive role a doctor's expertise plays in order to emit a positive diagnosis. However, when dealing with cytology samples in which cellular presence is inconclusive to discard either a positive or a negative call, an exhaustive analysis is required. Although it is uncommon, atypical glandular cells come up as a big challenge to specialists since most of these samples cause a bigger contribution to call false positives or false negatives. It will always be worse to call a false negative in patients rather than a false positive, for obvious reasons.

As discussed earlier, the contribution of this research focuses on the diagnosis of atypical cases of CN, through the study of atypical glandular cells (AGC). These types of cells present an atypical nucleus that exceeds the reactive or reparative changes but lacks endocervical adenocarcinoma in situ features. This expert system is expected to be a supporting tool in the decision-making process for cervical cancer diagnosis; the system is specially oriented to patients whose samples show AGC and as a consequence reduce false positives and false negatives within the diagnosis.

In Figure 1, cytology of atypical endocervical cells is shown, where the nuclei almost triple their normal size, and a suggestive feature for adenocarcinoma can be observed. Despite this, nucleus pigmentation and nucleolus presence (small spots inside nuclei) are only observed in normal or nonmaligns samples.

In Figure 2, a suggestive neoplasic sample is presented, which means it's likely to be a positive Cancer diagnosis due to its mostly dark coloring and cellular fragmentation. However, the grouping and size of nucleolus suggest a normal sample.

This research was carried out with the collaboration of the Cytopathology Department of Rio Blanco's Hospital, where the decision-making process for a patient's diagnose is made

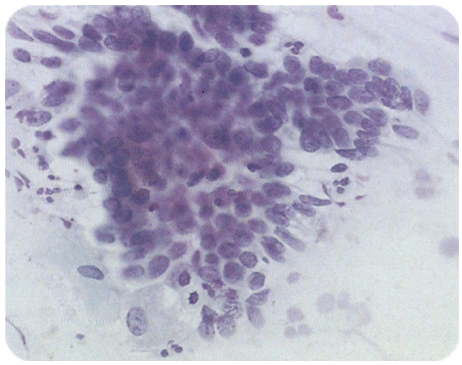


Figure 2: AGC, suggestive neoplastic.

under Mexico's General Hospital and NOM 014-SSA2 1994 procedures. These procedures consist of inspections of the sample by 2 different cytotechnologists that emit a diagnosis. If the cytotechnologists have different diagnosis, they discuss their point of views in order to reach an agreement and emit a diagnosis by consensus. In the first inspection (quick inspection done in less than one minute), all the present fields on the lamella are observed by a cytotechnologist, and the first diagnosis is emitted. Afterwards, in a slow inspection (at least four minutes), special emphasis is placed on the lamella fields that are not clear for injury rates. From this second inspection, a second diagnosis is also obtained. If a positive correlation exists between both diagnoses, a final diagnosis is emitted. Otherwise the sample goes through a consensus in order to get a final result. In the first two inspections, two cytotechnologists take part as examiners; however, in the debate session and consensus there are 6 other experienced cytotechnologists and 1 pathologist headed by Dr. José Antonio Palet Guzmán, Chief of the Department of Cytopathology. These people complete the diagnosis team at the HRRB.

A patients' treatment depends on if it is possible to decide on the presence of an AGC. For this reason, it is necessary to deeply analyze the cytological findings suggestive of benign changes and call them "negative for high grade epithelial Cancer injury" when possible. However, human eye interpretation is inconclusive in the cytological image analysis due to the different criteria on making a diagnosis of AGC. On the other hand, the different interpretations of a patient's clinical background will result in a consensus, since subjectivity appears on the scene at the moment of making a decision for a positive diagnosis of CN. This expert system allows unified criteria among many examiners as long as enough correct data is provided. This means that all the blanks in the interface must be filled by the user. If any error is given to the expert system, its reliability will decrease considerably.

2.2. Risk Factors for CN. The origin of cervical cancer is multifactorial. The most outstanding risk factor is infection of human papillomavirus (HPV). However, the patient's ages, and the early beginning of his sexual relations are factors to be considered extremely important [7].

Multiparity is another risk element to be considered. According to published studies, it has been proven that an immunologic and folic acid depression occurs. The more pregnancy numbers, the higher the risk of intraepithelial neoplasms [8]. In the same way, the age of the first gestation is of vital importance in CN diagnosis, since it helps to estimate the evolution of this disease. Having multiple sexual partners is another important factor, primarily if contraceptive methods are hormonal instead of barrier methods.

Some other important factors are the presence of sexually transmitted diseases (STD) and cervical injuries found in Pap smears (Pap smears is a test in which cells from the uterine cervix and endocervix are sampled, spread on a glass slide, stained, and interpreted by a cytotechnologist or pathologist based on the Bethesda system abnormal values. The PS is the "gold standard" method for early detection of HPV, herpes, trichomonad infections, CIN/dysplasia, and cervical CA accuracy PS interpretation is a subjective art based on experience of the screening cytotechnologist or pathologist. http://medical-dictionary.thefreedictionary.com/Pap+smear).

2.3. Detection and Diagnosis. There are cases that even though cytology is positive, a patient does not show malignant or precursor injuries. In other cases, cytology is negative, and in further checking, malignant anatomical changes are found which are not identified at first, incurring in false positives or false negatives [9].

The current suggested terminology for reporting results in cervical cytology is the Bethesda system. This system was developed for the National Cancer Institute (NCI) in 1988, with the purpose of providing a uniformed terminology to facilitate communication between the pathologist and gynecologist. The main purpose of this system is to inform the gynecologist the most information available to be used in a patients' treatment by the means of a descriptive report in which all the cytological aspects must be included (hormonal, morphologic, and microbiological levels) [10].

2.4. A Literature Review. As a part of this research, many other related works were consulted. These pieces of work are based on different technological and methodological techniques.

Deligdisch et al. present a classification of human papillomavirus (HPV) in teenagers. Here, they identify histological changes in the cervical uterus of infected teenagers with HPV and contrast their results with the results of older patients. The analysis presented was carried out by the means of image treatment techniques, and their final contribution was to assess cytopathic effects of HPV in teenagers [11].

Torres et al. assess the role of the main risk factors linked to high-risk squamos intraepithelial injuries in Cauca, Colombia. The methodology used in this research consists of statistic case studies and checkups. In addition, this study confirms the relation between HPV and the risk of cervix neoplasm. Data suggest that multiparity and exposure to carcinogens present in smoke from firewood are risk cofactors when HPV is presented [12].

Arcos' research results show that it was possible to classify 75 vaginal epithelial cells after using operations such as preprocessing, segmentation, and representation of digital images. Software was built which allows you to analyze cell classification parameters, such as relationship between nucleus cytoplasma, staining, saturation, luminance, and shape and position of the nucleus, with the purpose of identifying superficial, intermediate, basal, parabasal, and dysplastic cells [13].

In her research, Machorro describes the development of an expert system that eliminates subjectivity that may occur in the different viewpoints among colposcopy specialist. After this, the information is presented to gynecologists as an alternative support for making a decision in diagnosing the type of CN precursor injury based on fuzzy logics and colposcopic images processing [14].

Zarandi et al. explain a fuzzy expert system's performance for human brain tumors diagnosis by using T1 contrast magnetic resonance imaging. Fuzzy T2 image processing proposal has four different modules: preprocessing, segmentation, feature extraction, and approximate reasoning. The expert system was validated to prove its accuracy in real situations. The results showed that the system is superior in recognizing brain tumors and its grade [15].

In 2011, Tan presents a guide to facilitate the implementation of the Image Guided Brachytherapy (IGTB) which is one of the top-end techniques for cervix cancer treatment. This research was originally published by the Royal College of Radiologists (RCR) in 2009 [16].

Petric et al. ran a test in 13 patients in which high-risk-CTV (HR-CTV) was outlined by two observers in T and PT MR image plane, respecting the GYN GEC-ESTRO recommendations for 3D-image based cervix cancer brachytherapy. Contouring time was measured. HR-CTV sizes were compared, and conformity index (CI) was assessed. Inter observer variations in contour extent along eight radial directions were compared between delineation planes. After applying a standard treatment plan, an intercomparison between DVH-parameters V100, D90, and D100 for the HR-CTV was carried out [17].

Kande et al.'s, paper presents a novel approach for automated segmentation of the vasculature in retinal images. The approach uses the intensity information from red and green channels of the same retinal image to correct nonuniform illumination in color fundus images. Matched filtering is utilized to enhance the contrast of blood vessels against the background. The enhanced blood vessels are then segmented by employing spatially weighted fuzzy c-means clustering based thresholds which can well maintain the spatial structure of the vascular tree segments. The proposed method's performance is evaluated on publicly available DRIVE and STARE databases of manually labeled images. On the DRIVE and STARE databases, it achieves an area under the receiver operating characteristic curve of 0.9518 and 0.9602, respectively, being superior to those presented by state-of-the-art unsupervised approaches and comparable to those obtained with the supervised methods [18].

Shin et al. proposed the Index-Blocked Discrete Cosine Transform Filtering Method (IB-DCTFM) to design an ideal frequency range filter on DCT domain for biomedical signal which is frequently exposed to specific frequency noise such as motion artifacts and 50/60 Hz power line interference. IB-DCTFM removes unwanted frequency range signals on time domain by blocking specific DCT index on DCT domain. In simulation, electrocardiography, electromyography, and photoplethysmography are used as a signal source, and FIR, IIR, and adaptive filters are used for comparison with proposed IB-DCTFM. To evaluate filter performance, we calculated signal-to-noise ratio and correlation coefficient to a clean signal of each signal and filtering method, respectively. As a result of filter simulation, average signal to noise ration and correlation coefficient of IB-DCTFM are improved about 75.8 dB/0.477, and FIR, IIR, and adaptive filtering results are 24.8 dB/0.130, 54.3 dB/0.440, and 29.5 dB/0.200, respectively [19].

Rogel et al. proposed that French uterine cancer recordings in death certificates include 60% of "uterine cancer, Not Otherwise Specified (NOS)"; this hampers the estimation of mortalities from cervix and corpus uteri cancers. The purpose of this work was to study the reliability of uterine cancer recordings in death certificates using a case matching with cancer registries and estimated age-specific proportions of deaths from cervix and corpus uteri cancers among all uterine cancer deaths by a statistical approach that uses incidence and survival data [20].

Thomas proposed that external irradiation and brachytherapy are curable for the treatment of carcinoma of the cervix. The intention of radiotherapy is to optimize the irradiation of the target volume and to reduce the dose to critical organs. The use of imaging (computed tomography and magnetic resonance imaging added to clinical findings and standard guidelines) is studied in the treatment planning of external irradiation and brachytherapy in carcinoma of the cervix. Imaging allows for individualized and conformal treatment planning [9].

Zwahlen et al. purpose to determine the feasibility and benefits of optimized magnetic-resonance-imaging- (MRI-) guided brachytherapy (BT) for cancer of the cervix. A total of 20 patients with International Federation of Gynecology and Obstetrics Stage IB-IV cervical cancer had an MRI-compatible intrauterine BT applicator inserted after external beam radiotherapy. MRI scans were acquired, and the gross tumor volume at diagnosis and at BT, the high-risk (HR) and intermediate-risk clinical target volume (CTV), rectal, sigmoid, and bladder walls were delineated. Pulsed-dose-rate BT was planned and delivered in a conventional manner. Optimized MRI-based plans were developed and compared with the conventional plans. MRI-based BT for cervical cancer has the potential to optimize primary tumor dosimetry and reduce the dose to critical normal tissues, particularly in patients with small tumors [21].

Gordona et al.'s work is focused on the generation and utilization of a reliable ground truth (GT) segmentation for a large medical repository of digital cervicographic images (cervigrams) collected by the National Cancer Institute (NCI). The NCI invited twenty experts to manually segment a set of 939 cervigrams into regions of medical and anatomical interest. Based on this unique data, the objectives of the

current work are to (1) automatically generate a multiexpert GT segmentation map; (2) use the GT map to automatically assess the complexity of a given segmentation task; (3) use the GT map to evaluate the performance of an automated segmentation algorithm. The methods and conclusions presented in this work are general and can be applied to different images and segmentation tasks. Here they are applied to the cervigram database including a thorough analysis of the available data [22].

Xue et al. proposed that segmentation is a fundamental component of many medical image-processing applications, and it has long been recognized as a challenging problem. In this paper, they report our research and development efforts on analyzing and extracting clinically meaningful regions from uterine cervix images in a large database created for the study of cervical cancer. In addition to proposing new algorithms, they also focus on developing open source tools which are in synchrony with the research objectives. These efforts have resulted in three Web accessible tools which address three important and interrelated subtopics in medical image segmentation, respectively: the Boundary Marking Tool (BMT), Cervigram Segmentation Tool (CST), and Multi-Observer Segmentation Evaluation System (MOSES) [23].

As a result of the literature review, we can observe that, even if the uses of techniques of expert systems such as fuzzy logics and images interpretation are common in medicine, the study of atypical gland cells (AGC) is hardly explored.

In recent years, FL has been used in different types of machines, systems, and software, among other applications of everyday life. Some examples are lift controls, earthquake engineering, and even video games. However, among other applications that allow this technique, at the same time one of the most innovative and complex, is medicine. On the one hand, it is possible to take advantage of the imprecision tolerance, and on the other hand, to unify vital criteria for the decision making process. The main contribution of fuzzy logic relies on the easiness to deal with complex, hard to define, or lack of math-model problems that would allow for their solution. Research works such as that of Machorro (2010) show us how this technique lets us unify criteria and reduce subjectivity among different points of view of specialists on an issue.

Although image interpretation is considered as an "open field" in research, the progress in this area has not been by itself but has been mostly linked to other areas. The rise of digital image processing (DIP) can be seen in areas such as astrology, geology, electronics, and, most importantly for us: medicine. Image interpretation allows us to obtain data that the human eye is not able to perceive in CAT scans, X-rays, or cytological image, for cervix cancer detection.

Cancer cells research is nothing new, since the techniques mentioned above have made an incursion in the study of this disease since decades ago. For instance, we can talk about researches comprising since HPV as a risk factor in contracting cervix cancer by the means of statistical analysis as described in works of Deligdisch et al. and Torres et al. Some other researchers such as Arcos and Zarandi et al. have applied image interpretation to identify and classify cervical cells and cerebral tumors. On the other hand, works by Thomas, Zwahlen et al., Gordona et al., and Xue et al. focus on segmentation techniques for countless medical applications for image interpretation; for instance, regions of clinical interest extraction, MRIs guided by brachytherapy, or even optimized radiation of CTV (clinical target volume?), and reduction doses, on critic organs, among others.

However, a problem for experts on the matter is atypical cells analysis. In this case, atypical gland cells (AGC) are presented when diagnosing cervix cancer (consider revision in Spanish). The main contribution of this research is the study of AGC for their consideration within the process of cervix cancer diagnosis, and at the same time, the reduction of false positives and false negatives that come along with these types of diagnosis.

3. Development of the Expert System

This expert system was developed in 3 phases or stages (Figure 3). The first stage is the risk diagnosis; this stage consists of a fuzzy model that measures the risk for contracting CN according to the patient's clinical records. The second stage is the interpretation of a cytological image where a process and an image interpretation are carried out in order to obtain common parameters in images with regular diagnosis, cancer, and in squamous cell Atypical Glandular Cells (AGC). The third stage of the expert system links the obtained results in the fuzzy model of the first stage with the results of the images processing in order to create a second fuzzy model, in which injuries can be classified. This classification is given by three output variables: no alterations, atypical glandular cells (AGC), and malign positive.

The expert system uses Matlab's toolbox for the development of the fuzzy logic model and image interpretation. A graphic interface that integrates Matlab's tools (fuzzy logic and image interpretation) was created so that the expert system can be user-friendly.

3.1. Fuzzy Model of Risk Diagnose. In the decision-making process for CN diagnosis, risk factors are considered essential when an atypical glandular cell (AGC) occurs, since a diagnosis can encourage a debate among specialists, and this could eventually affect the clinical procedure followed by doctors that could lead to false positives or false negatives. As a result, the expert system starts with a fuzzy logics model that determines if a patient is at risk of developing CN according to their clinical records.

The fuzzy rules combine one or more fuzzy input sets (background) and establishe association with output fuzzy sets (consequence risk). This rule format is known as the Mamdani (Mamdani FLC (Fuzzy Logic Controler) proposed by Mamdani and Assilian in 1994. This controller uses the error $e(k)$ and the change of error $\Delta e(k)$ to produce changes in the output function of the controller.) type and works with a fuzzy controller that settles up a system on its work area.

As mentioned earlier, the variables to be used are obtained directly from the patient's clinical records and according to what specialists' consider significant to their diagnosis of CN. Identified variables and their sets are shown in Table 1.

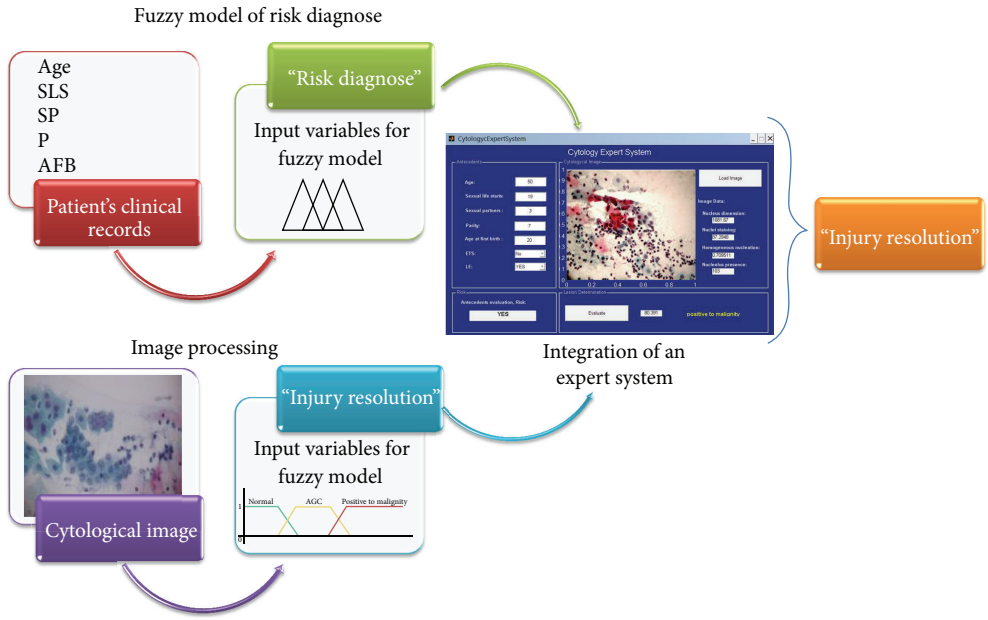


Figure 3: Integration of the expert system.

Table 1: Input variables of fuzzy model of risk diagnose.

Input variables	Label	Geometric shape	Interval
E (age)	Young	Trapezoid	(10 15 18 22)
	Adult	Trapezoid	(18 22 34 45)
	Mature	Trapezoid	(38 44 90 90)
IVSA (onset of sexual activity)	Young	Trapezoid	(10 15 18 22)
	Adult	Trapezoid	(18 22 34 45)
	Mature	Trapezoid	(38 44 90 90)
PS (sexual couples)	Few	Trapezoid	(0 2 3 4)
	Many	Trapezoid	(3 5 9 10)
	Unusual	Trapezoid	(9 15 20 20)
NG (number of pregnancies)	Few	Triangular	(0 0 0.5)
	Many	Triangular	(0.51 4 8)
		Trapezoid	(7 8 18 18)
PG (first pregnancy age)	Young	Trapezoid	(10 15 18 22)
	Adult	Trapezoid	(18 22 34 45)
	Mature	Trapezoid	(38 44 90 90)
ETS (sexually transmitted diseases)	No	Triangular	(0 0 0.5)
	Yes	Triangular	(0.51 1 1.5)
LE (cervical lesions)	No	Triangular	(0 0 0.5)
	Yes	Triangular	(0.51 1 1.5)
R (risk)	**Negative**	**Triangular**	**(0 0 0.5)**
	Positive	**Triangular**	**(0.51 1 1.5)**

The output variable for this model is considered with only two sets: positive risk and negative risk. These sets are given by independent triangle graphs; this means that a result will be negative if the gap is between 1 and 1.5, or positive if the output variable is superior to or equal to 1.51.

The modeling of these variables was made with the Matlab software. The intervals used for this model were provided by specialists based on their expertise. One example of this model is represented with the variable labeled "age" (Figure 4):

$$\mu_{\text{young}}(E) = \begin{cases} 0; & t < 10 \\ 1 - \dfrac{15 - t}{15 - 10}; & 10 \le t \le 15 \\ 1; & 15 \le t \le 18 \\ 1 - \dfrac{t - 18}{22 - 18}; & 18 \le t \le 22 \\ 0; & 22 > t, \end{cases}$$

$$\mu_{\text{adult}}(E) = \begin{cases} 0; & t < 18 \\ 1 - \dfrac{22 - t}{22 - 18}; & 18 \le t \le 22 \\ 1; & 22 \le t \le 34 \\ 1 - \dfrac{t - 34}{45 - 34}; & 34 \le t \le 45 \\ 0; & 45 > t, \end{cases} \quad (1)$$

$$\mu_{\text{mature}}(E) = \begin{cases} 0; & t < 38 \\ 1 - \dfrac{t - 40}{40 - 38}; & 38 \le t \le 40 \\ 1; & 40 > t. \end{cases}$$

The creation of the rules of inference in the system was determined based on the linguistic input variables and their

Table 2: Example of some inference rules.

Nucleus	E	IVSA	PS	NG	EPG	ETS	LE	Risk
1	Young	Young	Little	Null	Young	NO	No	Negative
2	Young	Young	Little	Null	Young	NO	Yes	Negative
3	Young	Young	Little	Null	Young	Yes	No	Negative
4	Young	Young	Little	Null	Young	Yes	Yes	Negative
5	Young	Young	Many	Null	Young	No	No	Positive
6	Young	Young	Many	Null	Young	No	Yes	Positive
7	Young	Young	Many	Null	Young	Yes	No	Positive
8	Young	Young	Many	Null	Young	Yes	Yes	Positive
9	Young	Young	Not common	Null	Young	No	No	Positive
10	Young	Young	Not common	Null	Young	No	Yes	Positive
11	Young	Young	Not common	Null	Young	Yes	No	Positive
12	Young	Young	Not common	Null	Young	Yes	Yes	Positive
13	Young	Young	Little	Little	Adult	No	No	Negative
14	Young	Young	Little	Little	Adult	No	Yes	Negative
15	Young	Young	Little	Little	Adult	Yes	No	Negative
16	Young	Young	Little	Little	Adult	Yes	Yes	Negative
17	Young	Young	Many	Little	Adult	No	No	Positive
18	Young	Young	Many	Little	Adult	No	Yes	Positive
19	Young	Young	Many	Little	Adult	Yes	No	Positive
20	Young	Young	Many	Little	Adult	Yes	Yes	Positive

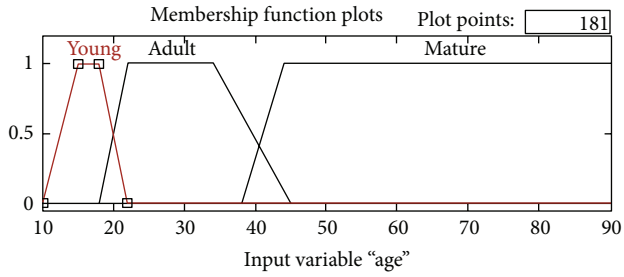


Figure 4: Modeling of variable "Age."

fuzzy sets. The combination of the seven variables along with their sets brought about 972 rules. After these inference rules were validated by specialists, only 360 rules were left to run the system. Some of these rules can be seen in Table 2.

The output value according to each reference rule was determined by the specialists.

The defuzzification process that uses the Matlab software calculates the center of gravity of images (centroid method) that are generated at the moment of decomposing each of the linguistic input variables; this means that this process analyzes the existing association between the membership grades of two joint sets and calculates the surface of each resulting image. In other words, the centroid method uses the classic form, in relation to the abscissa axis, the center gravity in rectangles is located in the middle of its base, and for triangles, it is found at the third part of its base according to the opposite side of the angle created by the hypotenuse and the base.

Finally, the total centroid is calculated. To obtain this variable, the addition of the surface product of each figure is divided by its centroid into the total surface.

3.2. Image Processing. Image processing consists of many different stages that make it possible to identify useful input parameters for an expert system enabling it to determine cancer precursor injuries. To provide an overview about image processing, Figure 5 is shown and explained.

Images processing consists of the use of techniques on one or more images in order to obtain data or specific features. Applied techniques in this research focus on the identification of endocervix cell nuclei, as well as their size, staining, shape, and interpretation. These features are further used as linguistic or input variables of the fuzzy logics model that make up the expert system of this research.

This study was based on cytological images. Cytological images are classified in three diagnoses: normal or negative to malignant, atypical glandular cells (AGC), and adenocarcinoma. When analyzing an image, specific features are expected to be found. These features or sets of features will define the sort of condition. In Figure 6, you can observe some of the important findings. Outlined in green you will find normal or negative to malign cells. Outlined in red, you can see metaplastic or evolving cells, which can turn into cancer. Outlined in blue, there are the inflammatory cells (also known as mast cells), which are not so relevant for analysis and diagnosis of an injury. Finally, in light orange you

Figure 5: Image processing.

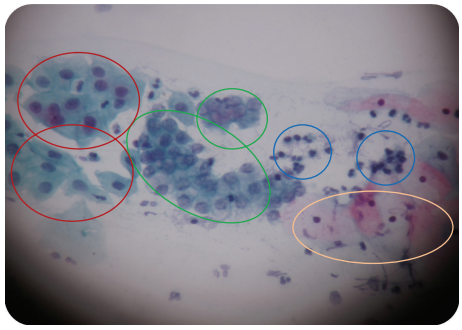


Figure 6: Benign cytological findings.

can observe exocervical cells that in this research we will not be looking at since these types of cells are studied in histology.

The findings which were discussed earlier are the characteristics that are intended to extract from each image. These features are considered by specialists as fundamental for CN diagnosis. The image-processing techniques applied in the Matlab software for this work will be discussed up next.

(a) Level Enhancement and Embossing. In this first stage of the process, it is intended to improve the color levels of the image by adjusting its brightness and contrast for the purpose of stressing the interest features in it. Applying any of these adjustments, you may observe noticeable changes in the images. In the next diagram, we can see the original image in the left part, just as a specialist would see through the microscope. On the other hand, on the right side we can see the same image with the first level adjustment applied. The image is acquired directly from the microscope through a photographic camera which contains low default levels of color, brightness, and contrast. For this reason, the levels must be adjusted in order to get a better quality of the image and outline the areas of interest. The levels of brightness and contrast were adjusted according to the specialist's experience and appreciation. Once the camera was ready according to the specialists' requests, the camera adjustments were set for the rest of the photographs.

In Figure 7, you can immediately identify the darkening of the nucleus and the main interest feature, as well as the cytoplasm definition. The background of the image could be mistaken with the cytoplasm when adjusting brightness and contrast levels, since this adjustment makes it clearer to identify. Moreover, the edge of the cell membrane is better defined for further analysis.

(b) Color Separation. Once the image has been adjusted, it is divided into three channels, or its color composition, this means, red, green, and blue channel, best known as RGB. In Figure 8, you can observe the channels composing the image.

The following step in the image processing is where nuclei are going to be identified. This step was decided to be carried out in the green channel by the specialists in the HRRB. The specialists based this decision on their expertise since this channel keeps and better marks the interest features.

The grayscale technique will be used to obtain some other parameters such as nuclei staining, dimension, and nucleus shape. In Figure 9, you can see the image converted into grayscale.

(c) Image Threshold. A threshold is a filter that works by dividing the image into two colors (black and white) and uses a middle spot or interest spot that can define a separation threshold in both colors according to the luminosity of the original image. If this is the case, threshold is used to divide the nucleus in an image, in order to obtain definite nuclei. In Figure 10, you will observe an image with a threshold over a green channel.

(d) Morphologic Operations. In image processing, morphology is a set of techniques based on shapes. This technique is applied to an input image (Figure 11) but does not affect its output. The set of techniques are dilation, erosion, closure, and disconnection.

(e) Labeling. This part of the image processing consists in identifying the number of objects that can be found in an image. Later, it is analyzed in order to obtain the area, perimeter, and centroid of the identified nuclei. This stage is performed in the green channel by using threshold and morphologic operations in order to determine the nucleus edges, which later will be used as a study object for parameter determination. In Figure 12, the process to identifying image nuclei can be observed.

By using a similar process to the prior, the number of nucleolus is obtained (Figure 13).

In order to deeply know the content of an image it is necessary to perform an analysis of the data gathered during its process. Furthermore, during this process it is possible to request data such as pixel size, nuclei diameters, symmetry, or eccentricity where specialists can observe features of interest.

One form of data interpretation in a clearer way is by using descriptive statistics. With this discipline we can obtain parameters that describe image information. For each data, images tend to settle in different ways; however, the images with the same condition tend to be alike. For this reason, after all the images from each condition are analyzed, the following variables are determined.

(i) *Nucleus dimension:* this refers to the size of each of the identified nucleus. To determine this parameter, the area of nucleus column is used and then you calculate

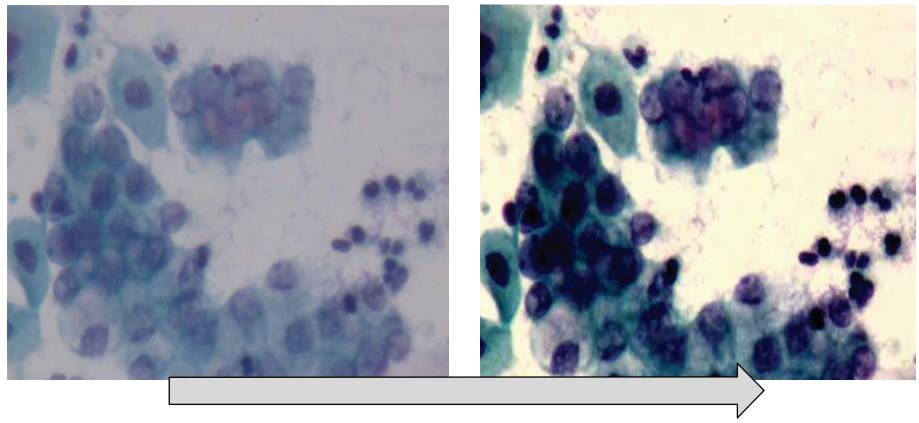


Figure 7: Level enhancement and embossing.

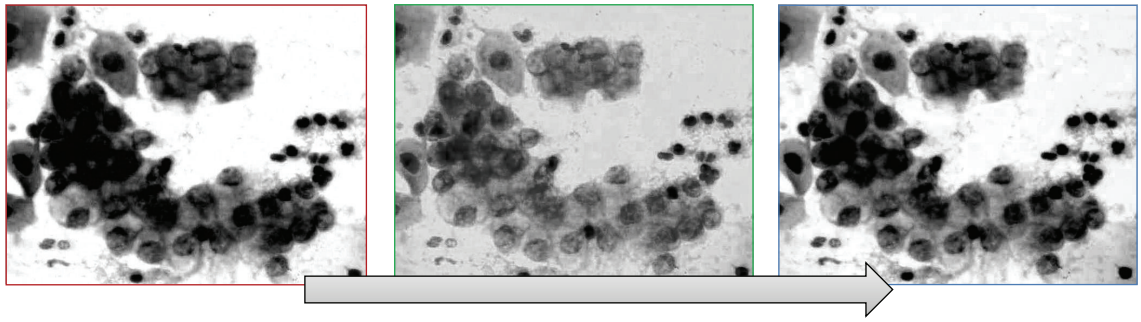


Figure 8: Color separation on RGB.

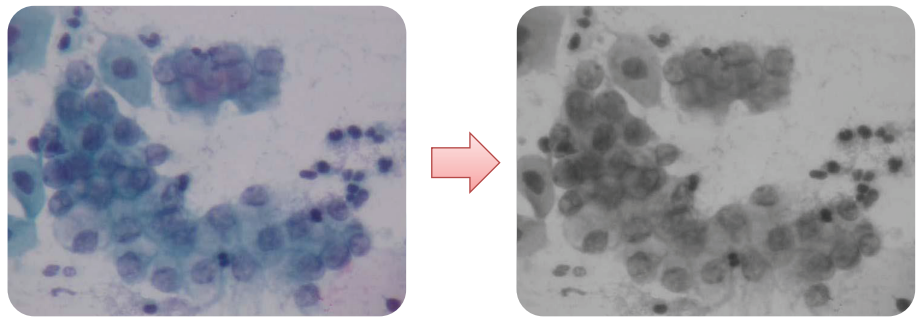


Figure 9: Image converted into grayscale.

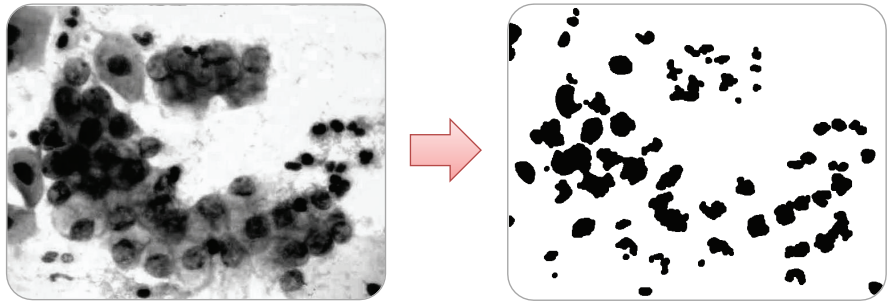
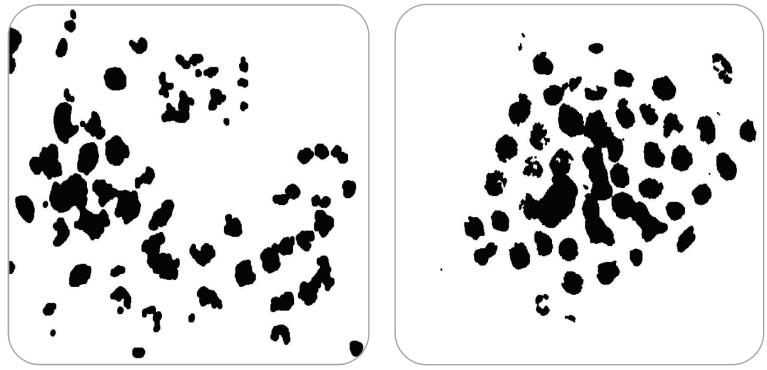
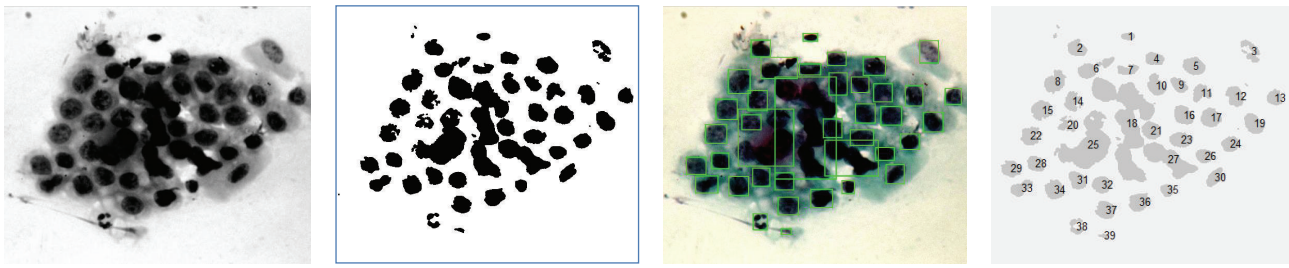


Figure 10: Threshold implementation for nuclei identification.



(a) (b)

Figure 11: Identification of nuclei.



(a) (b) (c) (d)

Figure 12: Identification and labeling of nuclei.

the average area of the nucleus that can be seen in the image:

$$DN = \frac{\sum \text{nucleus dimension (pixels)}}{\text{number of nucleus detected}}. \quad (2)$$

(ii) *Nuclei staining:* this refers to the pigmentation of the detected nuclei, that is to say how darkened the nuclei are, because the darker they are the more hyperchromatic they are considered. In order to obtain the nuclei staining appearing in an image, the nuclei pixel size column is used, and then the general average of the nuclei is calculated:

$$TN = \frac{\sum \text{pixel average of each nucleus}}{\text{number of nucleus detected}}. \quad (3)$$

(iii) *Homogeneous nucleation:* it is considered as homogeneous when identified nuclei are alike among them, meaning, when they are regular. The nuclei are considered to be regular when they tend to be rounded. In order to calculate homogeneous nucleation, the eccentricity column is used, and this column suggests that if eccentricity value in the nuclei tends to be 0 it is more rounded, but if the value tends to be 1 it is more elongated.

(iv) *Nucleolus presence:* one important part of the diagnostic criteria in CN is the ability to identify nucleolus in an image. Along with the image processing, a numeric data of the amount of identified nucleolus is obtained.

3.3. Integration of an Expert System: Fuzzy Model "Injury Resolution". An expert system is that which is capable of emulating a decision-making process involving a human being with inaccurate information. Also, it is capable of making a decision with incomplete information just the way a specialist in a subject would. The expert systems based on fuzzy logics are able to transform linguistic variables into numeric values by using fuzzy sets, in which each variable belongs in a certain degree to another Fuzzy set in a different degree. The basis of this knowledge is built on a specialist expertise through rules represented with phrases such as: if... then; if not... then, known as Inference Engine. In a similar way, just as the fuzzy model "Risk," the fuzzy model "Injury Resolution" was carried out within the Mamdani type.

The integration of the expert system during this research was carried out through a graphic interface in which the user only needs to type a patient's clinical background data such as age, beginning of sexual life, number of children, and so forth. Then, upload the cytological image sample to be analyzed and finally press the bottom "Evaluate." The output variable (risk) and the result from the image process constitute the

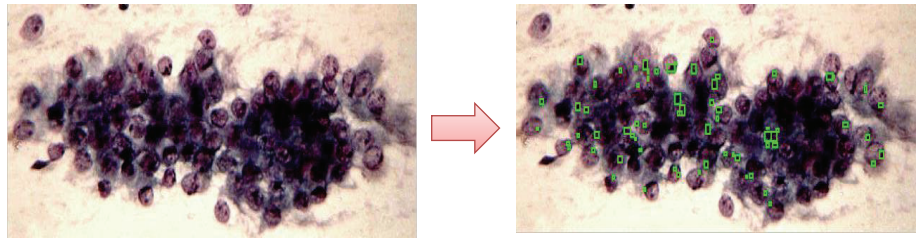


Figure 13: Identification of nucleolus.

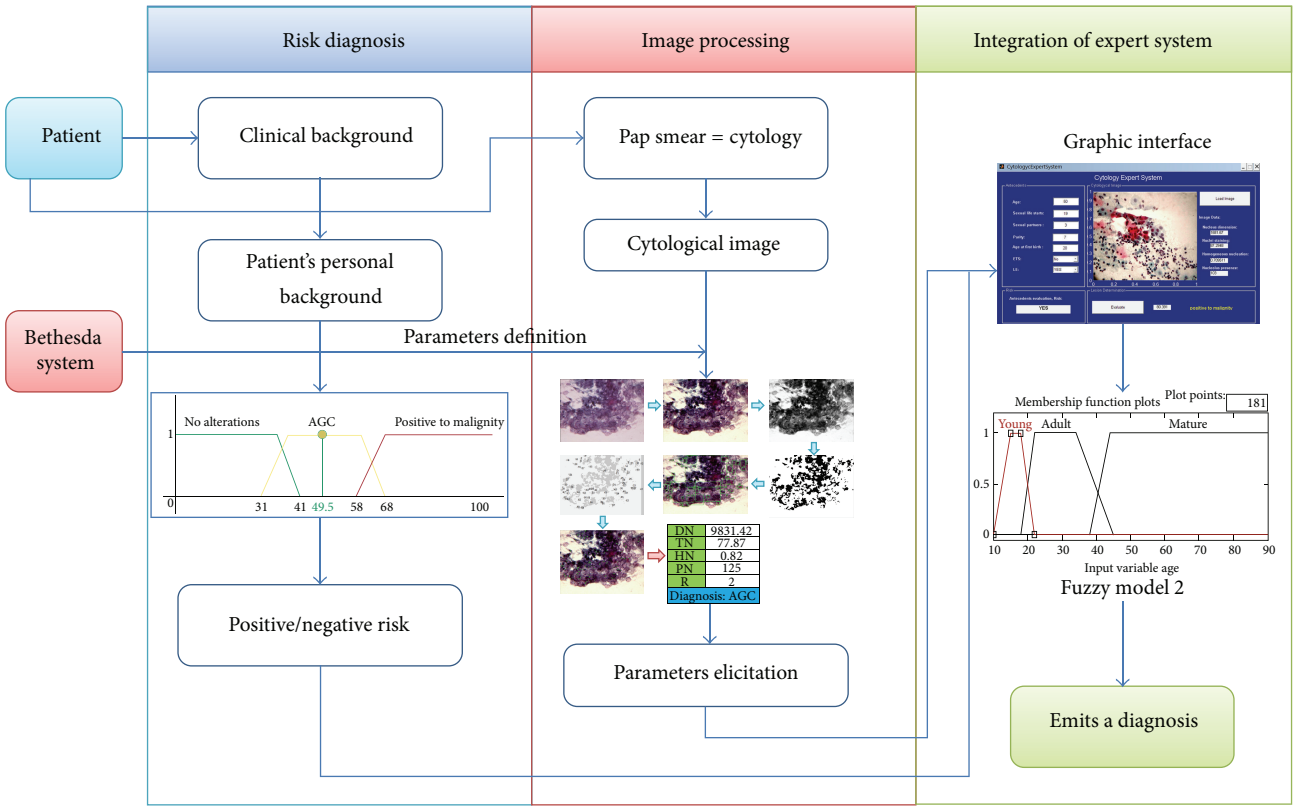


Figure 14: Expert Systems Process.

input variables of the second fuzzy model, which at the same time brings out the numeric and linguistic value that can suggest a diagnosis. The integration process can be observed in Figure 14.

The resulting data from the latter processing will provide parameters that will be turned into linguistic variables or input variables for the fuzzy model. The input variables to the fuzzy logics system are in Table 3.

As mentioned earlier, through the image processing it is possible to request data such as pixel size, nuclei diameters, symmetry, or eccentricity. Each of these data provides interest information to specialists since the analysis is made nucleus per nucleus (Figure 15).

One way to obtain parameters that describe an image's information is by Interval Estimation. Among images with the same condition, the data averages happen to be very similar. Due to the information provided by the image processing, theses intervals for the first four variables were set by using confidence intervals (Table 6). This means that the confidence interval of the average of each set would represent 100% of the set's membership in relation to the variable.

Each set for defuzzification of a variable was modeled in relation to the confidence intervals. Up next, the variable modeling "Nuclei Density" is presented (Figure 16), (4):

$$\mu_{\text{small}}(DN) = \begin{cases} 1; & 0 \le t \le 5000 \\ 1 - \dfrac{t - 5000}{6000 - 5000}; & 5000 \le t \le 6000 \\ 0; & t \ge 600, \end{cases}$$

Table 3: Input variables "Inyury resolution (DL)."[1]

Input variables	Label	Geometric shape	Interval
Nucleus dimensión (DN)	Small	Trapezoid	(0 0 5000 6000)
	Medium	Trapezoid	(4500 6800 9000 10000)
	Large	Trapezoid	(9000 9900 25000 25000)
Nuclei staining (TN)	Clear	Trapezoid	(50 50 80 82)
	Media	Trapezoid	(79 82 87 92)
	Dark	Trapezoid	(87 92 150 150)
Homogeneous nucleation (HN)	Regular	Trapezoid	(0.5 0.5 0.74 0.77)
	Irregular	Trapezoid	(0.75 0.8 1 1)
Nucleolus presence (PN)	Null	Triangular	(0 0 2)
	Few	Trapezoid	(2 9 20 30)
	Many	Trapezoid	(24 40 150 150)
Risk (R)	Negative	Triangular	(0.5 1 1.5)
	Positive	Triangular	(1.51 2 2.5)
(DL) Injury resolution	Normal	Trapezoid	(0 0 31 41)
	AGC	Trapezoid	(31 41 58 68)
	Positive to malignity	Trapezoid	(58 68 100 100)

[1]The initials presented in each one of input and output variables come from the translation from Spanish since users are native Spanish speakers. The interface of the expert system is, for now, only available in Spanish.

$$\mu_{\text{medium}}(DN) = \begin{cases} 0; & t < 4500 \\ 1 - \dfrac{6800 - t}{6800 - 4500}; & 4500 \le t \le 6800 \\ 1; & 6800 \le t \le 9000 \\ 1 - \dfrac{t - 9000}{10000 - 9000}; & 9000 \le t \le 10000 \\ 0; & t \ge 10000, \end{cases}$$

$$\mu_{\text{large}}(DN) = \begin{cases} 0; & t \le 9000 \\ 1 - \dfrac{9900 - t}{9900 - 9000}; & 9000 \le t \le 9900 \\ 1; & 9900 \le t \ge 25000. \end{cases} \tag{4}$$

One hundred and eight inference rules were created (5) which were validated by specialist doctors with the purpose of determining that all of them were feasible in relation to the variables' combination; for example, if in an image a "dark" nucleus is found it is not possible to detect the nucleolus; therefore, this rule turns out invalid:

$$DM(3) \times TN(2) \times HN(2) \times PN(3) \times R(2) = 108 \text{ Inference Rules.} \tag{5}$$

Some of these rules can be seen in Table 4.

4. Test and Results

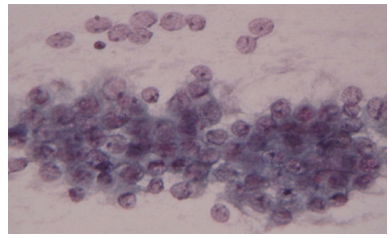
A graphic interface allows interaction with users in an easier way within the system. The expert system interface presented here (Figure 16) includes a background section (upper left), in which numeric and linguistic data is typed. Then at the upper right, an "upload image" button is included; here the image is selected from stored images from which study cases are generated. The background evaluation box (lower left) allows one to know if the patient has or does not have risk of contracting CN according to the clinical analysis. This information can be consulted by pressing the "evaluate" button in the "Injury Resolution" section. It is in this section that the numeric and linguistic complete analysis value of a case will appear (background + cytological image), and in the same way, numeric values for each parameter of the image processing (image data) will be shown.

The numeric value of the image processing is proportional to the malignancy grade of each case; however, the injury grade of each sample such as cervical intraepitelial neoplasm (CIN) I, II, and III was defined by specialists. This means that the system is capable of offering a diagnosis to cases with no variation (NV), atypical glandular cells (AGC), and positives to malignity, which are considered as the basis for CN diagnosis. These positive cases can be classified as infiltrating, invasive, adenocarcinoma, and so forth according to time and gravity of the injury. If an injury reoccurs or has not been treated before, specialists' expertise intervene.

The system tested the cases received at the hospital from 2010 to 2012. Given the fact that AGC are not common, these cases were recorded by the hospital and labeled as "suspicious" of AGC. It is important to mention that only a few patients present cases where AGC are involved.

In the first model validation of the fuzzy system (risk diagnosis), 20 gynecologic cases of patients from Rio Blanco's Regional Hospital (HRRB in Spanish) were considered, since these cases already had a diagnosis. These cases were the population sampling. The criteria for choosing these cases was that they had a cytological diagnosis, which was no longer than 1 year, and at the same time, their lamella was well preserved for a later photograph.

Results of image analysis-CITO 15c normal										
Nucleus	DN	TN	HN	Deviation	Asymetry	Kurtosis	Metrics	Relcel	Pixel mayor	Rel pixel
1	720.00	141.37	0.94	13.29	1.35	1.57	0.27	0.63	140.15	0.34
2	1377.00	135.94	0.74	26.75	−0.40	−1.31	0.41	0.75	95.23	0.18
3	2039.00	134.18	0.72	23.54	1.18	1.52	0.54	0.73	141.04	0.18
4	736.00	119.26	0.94	11.05	−0.03	−1.34	0.40	0.77	104.46	0.16
5	859.00	129.24	0.90	16.74	−0.22	−1.54	0.40	0.88	148.08	0.26
6	784.00	143.14	0.77	18.96	1.80	2.93	0.49	0.23	135.38	0.30
7	592.00	100.99	0.80	16.57	0.05	−1.50	0.55	0.76	79.36	0.17
8	523.00	117.76	0.79	20.87	0.87	−0.25	0.54	0.59	97.41	0.24
9	1011.00	126.77	0.95	8.59	0.20	−0.43	0.31	0.77	130.50	0.23
10	562.00	113.76	0.91	26.55	0.26	−1.23	0.47	0.95	80.86	0.17
11	1814.00	92.39	0.67	40.29	0.56	−0.72	0.58	0.95	47.38	0.23
12	1611.00	109.28	0.75	22.71	−0.22	−1.38	0.40	0.96	135.58	0.17
48	626.00	45.94	0.74	17.76	0.67	−0.42	0.55	0.94	26.00	0.20
49	7871.00	70.80	0.65	29.24	0.58	−0.68	0.37	1.00	38.57	0.26
50	3824.00	52.02	0.58	12.20	−0.59	−0.85	0.39	1.00	61.00	0.25
51	2051.00	90.75	0.88	20.55	0.96	0.78	0.34	0.96	77.07	0.19
52	11587.00	68.00	0.89	19.45	0.29	−1.41	0.23	1.00	48.72	0.18
53	620.00	81.94	0.77	22.47	1.35	1.29	0.52	0.90	77.14	0.24
54	1480.00	28.21	0.72	17.72	1.01	−0.26	0.20	0.98	15.62	0.25
55	3268.00	115.65	0.83	17.68	0.19	−0.71	0.36	0.90	110.71	0.24
56	4057.00	98.76	0.43	29.15	0.74	−0.41	0.54	0.90	89.75	0.21
57	4217.00	115.36	0.68	21.80	0.38	−0.82	0.41	0.89	100.50	0.19
58	4093.00	83.18	0.80	25.46	−0.07	−0.75	0.27	1.00	91.53	0.15
59	5583.00	59.61	0.66	19.44	0.88	−0.38	0.41	0.99	43.39	0.21
60	6427.00	74.61	0.58	13.59	0.46	−1.36	0.42	1.00	62.10	0.17
61	3449.00	140.48	0.70	18.99	0.89	0.57	0.43	0.89	132.75	0.24
62	923.00	100.12	0.93	9.86	1.10	1.14	0.42	1.00	88.75	0.21
31.50	**6548.16**	**81.65**	**0.78**	17.55	0.39	−0.49	0.39	0.91	73.50	0.21



Summary				
Nucleus	DN	TN	HN	PN
31.50	**6548.16**	**81.65**	**0.78**	97

Figure 15: Image processing results.

Table 4: Some examples of inference rules.

Rule number	Nucleus dimension	Nucleus tintion	Nucleus homogeneity	Micronucleus presence	Risk	Lesion
1	Small	Light	Regular	Null	Negative	Normal
2	Small	Medium	Regular	Null	Negative	Normal
3	Small	Dark	Regular	Null	Negative	Normal
4	Small	Light	Regular	Few	Negative	Normal
5	Small	Medium	Regular	Few	Negative	Normal

These cases were selected by the specialists because in 2011 a new microscope with photograph camera was acquired by the hospital, and therefore, the most recent lamella and complete data record cases were chosen.

Note: not all of the cases were positive to AGC, since the first diagnosis was just "suspicious for AGC." The latter was clarified further in the decision-making process named "session" where some cases where dropped as AGC. During the running test of the expert system, the final diagnosis emitted in the session at the HRRB was taken into consideration.

The system hypothesis is as follows.

According to the clinical backgrounds of a patient, if she is at risk of developing Cervical Cancer, the result of the cytology diagnose must be positive, whereas, if the result of the system shows that

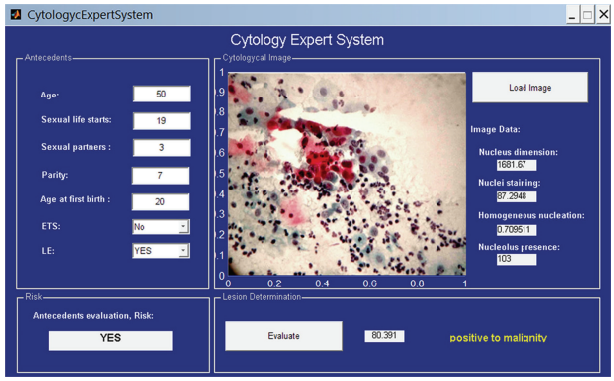


Figure 16: Expert system interface.

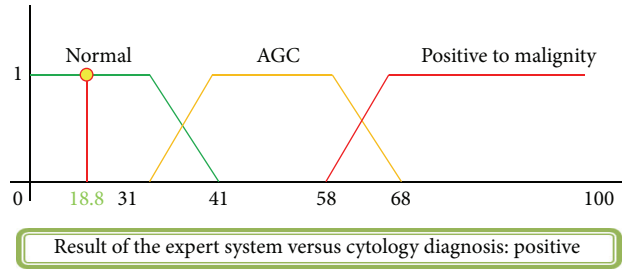


Figure 17: Injury resolution.

Table 5: Final results of expert system.

	Final results		
Case	Cytology diagnosis	Result of expert system	Correlation
1	Normal	Normal	Positive
2	Normal	Normal	Positive
3	Normal	Normal	Positive
4	Normal	Normal	Positive
5	AGC	AGC	Positive
6	Adenocarcinoma	Adenocarcinoma	Positive
7	AGC	AGC	Positive
8	Adenocarcinoma	Adenocarcinoma	Positive
9	Normal	Normal	Positive
10	Adenocarcinoma	Adenocarcinoma	Positive

there is no risk in developing Cervical Cancer, the diagnosis of the patient must be negative to cancer in relation to its last cytology.

As a result at the end of the validation of the system, we had a 100% correlation between the fuzzy model "Risk" and the cytology diagnosis emitted by the specialists. The system is valid to determine if a patient has a risk of developing CN only by looking into her clinical background.

The final system was tested with 10 diagnosed cases because they were the only ones that had a complete file at the moment of the test run. Another criterion was that since these cases were atypical there were not many of these kinds available. These 10 cases selected by the Head of Pathology were cases that were not able to be diagnosed at first and it was necessary to examine them again. Once the second tests were made and diagnoses were emitted, these cases became immediate candidates to run the expert system. The period in which these cases were taken for studying and running of this research was from May 2010 to January 2012.

The presence of risk for these cases was assessed with the first fuzzy model "Risk," in such a way that, in order to evaluate the model "Injury Resolution", a different output value was used in relation to the first model.

Up next, 3 out of the 10 cases tested are presented by the expert system. All of the cases were treated as AGC when examined for the first time by the specialists in the cytology sample, whereas their clinical backgrounds did not show any risk of contracting CN.

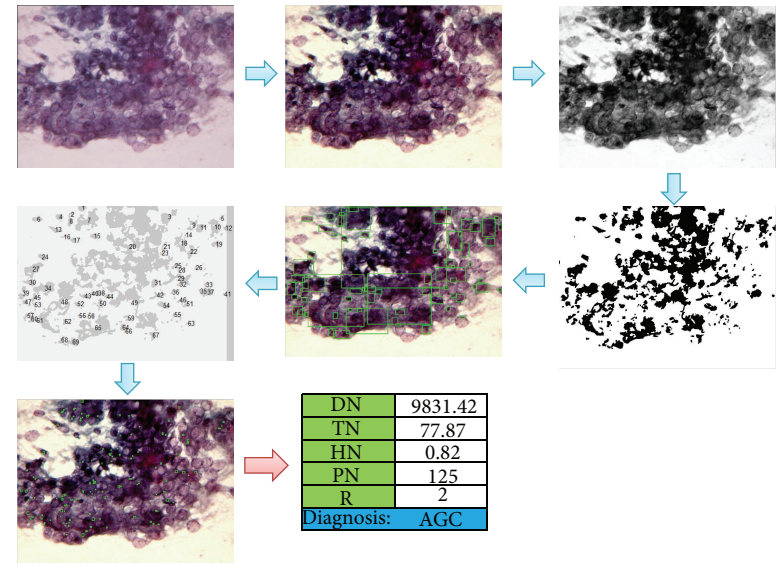
4.1. Case 1: Normal Cytology, Negative to Malignity. A 47-year-old woman, beginning her first sexual relation at the age of 23, 2 sexual partners and 1 gestation completed at the age of 27. No previous sexual transmission disease (STD) recorded. No cervical injuries found during exploration. Cytology diagnosis provided by the specialist: Negative to malignity = No alterations. The result of the system is shown in Figure 17.

4.2. Case 5: Cytology, Positive to Malignity, AGC. A 43-year-old woman, beginning her first sexual relation at the age of 16, 1 sexual partner and 7 gestations completed at the age of 28. No previous sexual transmission disease (STD) recorded. No cervical injuries found during exploration. Cytology diagnosis provided by the specialist: AGC. The result of the system is shown in Figure 18.

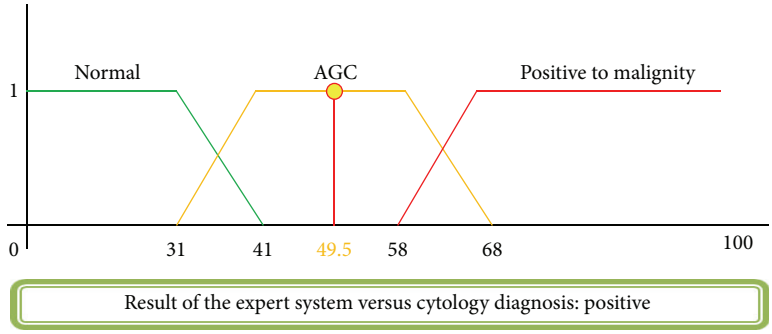
4.3. Case 8: Positive Cytology, Adenocarcinoma. A 50-year-old woman with first sexual relation at the age of 19, 3 sexual partners, and 7 gestations completing the first at the age of 20 via vaginal channel. No STD recorded. Cervical injuries found during exploration. Cytology diagnosis provided by the specialist: Adenocarcinoma = precursor injury of CN. The result of the system is shown in Figure 19.

In Table 5, the obtained results from the 10 diagnosed cases in HRRB are shown. Although it is true that the results obtained from the tests cannot show a significant reduction of false positives or false negatives, in these cases, 100% correlation between the doctors and the system was reached; however, a further test with more cases is expected. It is expected that with the implementation of this tool the percentage in the reduction of false positives and false negatives can be quantified.

The expert system is a supportive tool in the cervical cancer diagnoses at the HRRB, but the final decision is made by the specialist. However, in this section it is proven that tools and techniques such as image processing and fuzzy logics are as capable of emulating a decision-making process as much as a specialist. The image processing turns out to be a very interesting tool since it is through these kinds of techniques that detailed and accurate information can



(a)



(b)

Figure 18: Injury resolution.

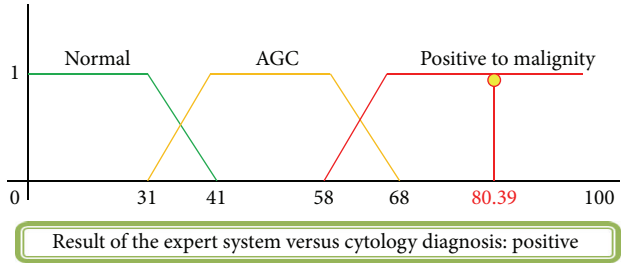


Figure 19: Injury resolution.

be consulted rather than just the subjectivity of the human eye in manual diagnosing. At the same time that fuzzy logics is a helpful tool for meeting criteria, it also helps model knowledge and expertise from specialists. Finally, the implementation of interfaces helps users to understand and interact in a simpler and easier way with the expert system.

As far as the data is reliable, the system will be able to make a diagnostic suggestion which helps the specialist to offer patients an opportune and appropriate treatment.

5. Conclusions

Cervix neoplasia (CN) is a public health matter; it is one of the most important causes of death from neoplasia among female population. Nevertheless, this type of cancer is preventable and its treatment can be considered relatively easy, when the diagnosis is opportune.

The Pap smear is not a diagnostic test, it is just a screening checkup that differentiates from those women who could

Table 6: Confidence interval DN.

	Small	Medium	Large
Data			
Sample size:	5	5	8
Sample average:	3,536.71	7,658.24	13,747.13
Population variance:	1,611,045.74	2,283,910.70	17,225,203.65
Confidence level:	0.99	0.99	0.99
Results			
Alfa (α):	0.01	0.01	0.01
$Z_{\alpha/2}$ =	2.58	2.58	2.58
Z_{α} =	2.33	2.33	2.33
	μ interval, σ known	μ interval, σ known	μ interval, σ known
	2,074.58 < μ < 4,998.84	5,917.35 < μ < 9,399.13	9,967.45 < μ < 17,526.80
	1,402.24 4,658.66	6,045.07 9,831.42	10,080.30 22,828.24

have cervical injuries from those who cannot. The results from these tests are not always "accurate." In many occasions, cytology can be positive, but at the end, the patient may not have cancer precursor or malignant injuries; whereas in other occasions, even though cytology is negative there may be malignant changes that were not first detected. Unfortunately, although this situation is very serious, it is also very common due to the presence of atypical cases in which samples with insufficient features to be diagnosed with CN tend to malignity and cannot be diagnosed as negative to malignancy (atypical gland cells) either. Human eye perception turns out insufficient for the decision making in the diagnosis. This is why a tool that helps doctors perform a more accurate test is of vital importance. Expert systems can be very useful in the decision making for atypical cases.

An expert system based on image processing and fuzzy logics applied on this type of illness allows a less subjective decision-making process, meeting existing diagnostic criteria in order to have more accurate information of the studied image. The implementation of the two techniques in this work that integrate the expert system are not feasible in isolation, since the parameters obtained through the image processing are not enough to emit a diagnosis. These results may vary from one to another; this means that it is fuzzy logics that integrates these techniques just the way a specialist would.

The expert system developed in this research has shown a great effectiveness in predicting if a patient could tend to CN according to her background or risk factors. However, there are many incomplete files that affect the efficiency of the system, since it depends on the input data. Knowledge integration from different fields is indispensable for the development of systems capable to emulate the decision-making process of an expert, since this innovates and enriches expert systems.

In the image processing there is a need to take in some considerations at the moment of taking photographs; this means, if the sample o lamella which is going to be studied is polluted, the photograph's quality will be doubtful, and emitted results from the processing may turn out nonvalid or uncertain. The integration of a first model and the image processing as information sources result in a more efficient expert system. Despite the fact that the system shows 100% positive correlation in relation to the specialist, it is necessary to carry out more tests and determine the error percentage that this might have.

In conclusion, the work presented here is considered as a support tool for specialists and is not intended to replace their invaluable labor. Criteria and considerations for each case can be known only by specialists and only they can make decisions based on their expertise. On the other hand, the implementation of an expert system that only requires a computer and software of commercial use makes it a very feasible tool for areas of low resources or marginalized zones. The implementation of the expert system is easy handling for doctors, since the graphic interface turns out to be user-friendly. The system can also be used as a training tool for new doctors, since within it, information, knowledge, and expertise from certified specialists are stored. Finally, it is worth mentioning that the contribution of this research relies on the support in the decision-making process of CN diagnosis. With this, it is expected to reduce the number of false positives and false negatives in cases where atypical cases occur.

References

[1] OMS Cáncer, Diciembre, 2011 http://www.who.int/mediacentre/factsheets/fs297/es/index.html/.

[2] D. Solomon and R. Nayar, "El Sistema de Bethesda para informar la citología cervical," *Buenos Aires*, vol. 37, no. 3, 2005.

[3] http://onlinelibrary.wiley.com/doi/10.3322/caac.20107/pdf.

[4] F. Serman, *Cancer Cervicouterino: Epidemiologia, Historia Natural y rol del Virus Papiloma Humano: Perspectivas en prevención y tratamiento*, Revista chilena de obstetricia y ginecología, 2002.

[5] SSA, *Programa De Acción*, Cáncer cérvico uterino, 2006.

[6] F. Meneses-González, M. Cos-Arroyo, and R. Tapia-Coger, *Prevalencia de uso de la Prueba de Papanicolau en Mujeres de 15 a 49 Años en México*, Revista Instituto Nacional de Cancerología, Mexico, México, 1999.

[7] G. B. Kande, P. V. Subbaiah, and T. S. Savithri, "Unsupervised fuzzy based vessel segmentation in pathological digital fundus

images," *Journal of Medical Systems*, vol. 34, no. 5, pp. 849–858, 2010.

[8] H. S. Shin, C. Lee, and M. Lee, "Ideal filtering approach on DCT domain for biomedical signals: index blocked DCT filtering method (IB-DCTFM)," *Journal of Medical Systems*, vol. 34, no. 4, pp. 741–753, 2010.

[9] L. Thomas, "Apport de l'imagerie dans la planification radiothérapique des cancers du col utérin," *Cancer/ Radiothérapie*, vol. 4, no. 2, pp. 127–132, 2000.

[10] NOM 014-SSA2, para la prevención, detección, diagnóstico, tratamiento, control y vigilancia, 1994.

[11] L. Deligdisch, C. R. De Resende Miranda, H. S. Wu, and J. Gil, "Human papillomavirus-related cervical lesions in adolescents: a histologic and morphometric study," *Gynecologic Oncology*, vol. 89, no. 1, pp. 52–59, 2003.

[12] C. H. S. Torres, M. P. A. Aragón, and L. O. Aristizabal, *Papilomavirus y Factores asociados a Neoplasm Intraepitelial Cervical de Alto Grado en Cauca, Colombia*, Revista de Salud Pública, Bogotá, Colombia, 2006.

[13] K. M. Arcos, "Identificacion de celulas displasicas en el epitelio vaginal," *Umbral Científico*, no. 9, pp. 104–114, 2006.

[14] F. R. Machorro, *Identificación de Lesiones de Cáncer Cérvico Uterino Mediante Tratamiento de Imágenes y Lógica Difusa*, Tesis de grado. Instituto Tecnológico de Orizaba, 2010.

[15] M. H. F. Zarandi, M. Zarinbal, and M. Izadi, "Systematic image processing for diagnosing brain tumors: a Type-II fuzzy expert system approach," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 285–294, 2011.

[16] L. T. Tan, "Implementation of image-guided brachytherapy for cervix cancer in the UK: progress update," *Journal of Medical Systems*, vol. 23, no. 10, pp. 681–684, 2011.

[17] P. Petric, J. Dimopoulos, C. Kirisits, D. Berger, R. Hudej, and R. Pötter, "Inter- and intraobserver variation in HR-CTV contouring: intercomparison of transverse and paratransverse image orientation in 3D-MRI assisted cervix cancer brachytherapy," *Radiotherapy and Oncology*, vol. 89, no. 2, pp. 164–171, 2008.

[18] G. B. Kande, P. V. Subbaiah, and T. S. Savithri, "Unsupervised fuzzy based vessel segmentation in pathological digital fundus images," *Journal of Medical Systems*, vol. 34, no. 5, pp. 849–858, 2010.

[19] H. S. Shin, C. Lee, and M. Lee, "Ideal filtering approach on DCT domain for biomedical signals: index blocked DCT filtering method (IB-DCTFM)," *Journal of Medical Systems*, vol. 34, no. 4, pp. 741–753, 2010.

[20] A. Rogel, A. Belot, F. Suzan et al., "Reliability of recording uterine cancer in death certification in France and age-specific proportions of deaths from cervix and corpus uteri," *Cancer Epidemiology*, vol. 35, no. 3, pp. 243–249, 2011.

[21] D. Zwahlen, J. Jezioranski, P. Chan et al., "Magnetic resonance imaging-guided intracavitary brachytherapy for cancer of the cervix," *International Journal of Radiation Oncology Biology Physics*, vol. 74, no. 4, pp. 1157–1164, 2009.

[22] S. Gordona, S. Lotenberga, R. Longb, S. Antanib, J. Jeronimoc, and H. Greenspana, "Evaluation of uterine cervix segmentations using ground truth from multiple experts," *Journal of Medical Systems*, vol. 33, no. 3, pp. 205–216, 2009.

[23] Z. Xue, L. R. Long, S. Antani, L. Neve, Y. Zhu, and G. R. Thoma, "A unified set of analysis tools for uterine cervix image segmentation," *Computerized Medical Imaging and Graphics*, vol. 34, no. 8, pp. 593–604, 2010.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 373401, 11 pages
http://dx.doi.org/10.1155/2013/373401

Research Article

Detection of Structural Changes in Tachogram Series for the Diagnosis of Atrial Fibrillation Events

Francesca Ieva, Anna Maria Paganoni, and Paolo Zanini

Dipartimento di Matematica, Modellistica e Calcolo Scientifico (MOX), Politecnico di Milano, Via Bonardi 9, 20133 Milano, Italy

Correspondence should be addressed to Anna Maria Paganoni; anna.paganoni@polimi.it

Received 25 October 2012; Accepted 25 March 2013

Academic Editor: Linamara Rizzo Battistella

Atrial Fibrillation (AF) is the most common cardiac arrhythmia. It naturally tends to become a chronic condition, and chronic Atrial Fibrillation leads to an increase in the risk of death. The study of the electrocardiographic signal, and in particular of the tachogram series, is a usual and effective way to investigate the presence of Atrial Fibrillation and to detect when a single event starts and ends. This work presents a new statistical method to deal with the identification of Atrial Fibrillation events, based on the order identification of the ARIMA models used for describing the RR time series that characterize the different phases of AF (pre-, during, and post-AF). A simulation study is carried out in order to assess the performance of the proposed method. Moreover, an application to real data concerning patients affected by Atrial Fibrillation is presented and discussed. Since the proposed method looks at structural changes of ARIMA models fitted on the RR time series for the AF event with respect to the pre- and post-AF phases, it is able to identify starting and ending points of an AF event even when AF follows or comes before irregular heartbeat time slots.

1. Introduction

During the last 20 years, there has been a widespread interest in the study of variations in the beat-to-beat timing of the heart, known as Heart Rate Variability (HRV) [1, 2]. This is due to several different reasons. HRV has been reported as strong predictor of cardiovascular mortality, and it is one of most popular parameter to assess the autonomic tone (see [3] and the references therein for a detailed discussion). Moreover, it represents a noninvasive way to assess post-surgical risks (see, e.g., [4]) or to investigate and tune gold standard practices [5]. Nevertheless, as highlighted in [6], the potential for HRV to be used widely in clinical practice remains to be established. When the sinus rhythm is normal, the tachogram series (i.e., the series of RR intervals; see Figure 1(a)) presents spontaneous beat-to-beat oscillations related to the autonomic nervous system regulatory action [7]. On the other hand, during arrhythmias, the spontaneous RR variability is perturbed and the spectral pattern changes according to the generating mechanisms of arrhythmia (see [8, 9]). Atrial Fibrillation (AF) is the most common cardiac arrhythmia, and involves the two upper chambers (atria) of the heart [10]. During AF, the normal electrical impulses generated by sinoatrial node are overwhelmed by disorganized electrical impulses that originate in the atria and pulmonary veins, leading to conduction of irregular impulses to the ventricles that generate the heartbeat. The result is an irregular heartbeat (see Figure 1(b)), which may occur in episodes lasting from minutes to weeks, or it could occur all the time for years. The natural tendency of AF is to become a chronic condition, and chronic AF leads to an increase in the risk of death.

The main device used in order to investigate the heartbeat is the Electrocardiogram (ECG) [11]. This diagnostic tool measures and records the electrical activity of the heart in details. The interpretation of these details allows for diagnosis of a wide range of heart diseases and AF among others. A stylized shape of an ECG is depicted in Figure 2. In general, atrial contraction shows up as the P wave; ventricular contraction is identified as a series of three waves, Q, R, and S, known as the QRS complex. The third wave in an ECG is the T wave. It reflects the electrical activity produced

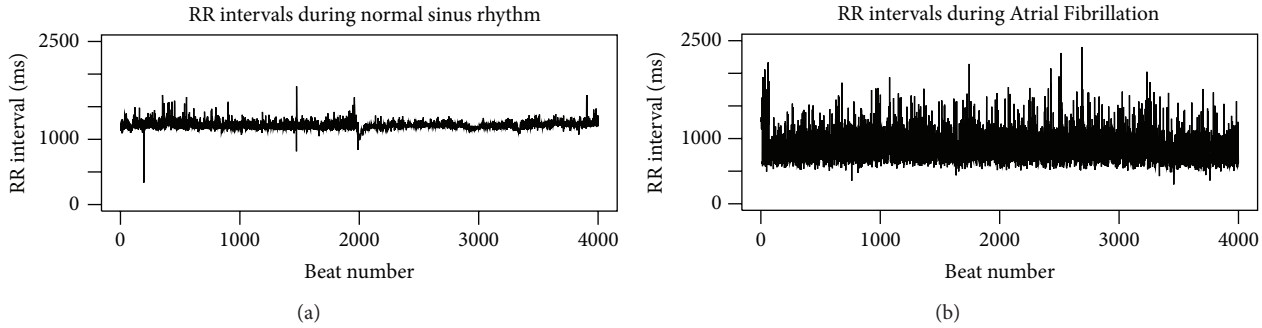


Figure 1: Typical series of RR intervals during normal sinus rhythm (a) and during Atrial Fibrillation (b).

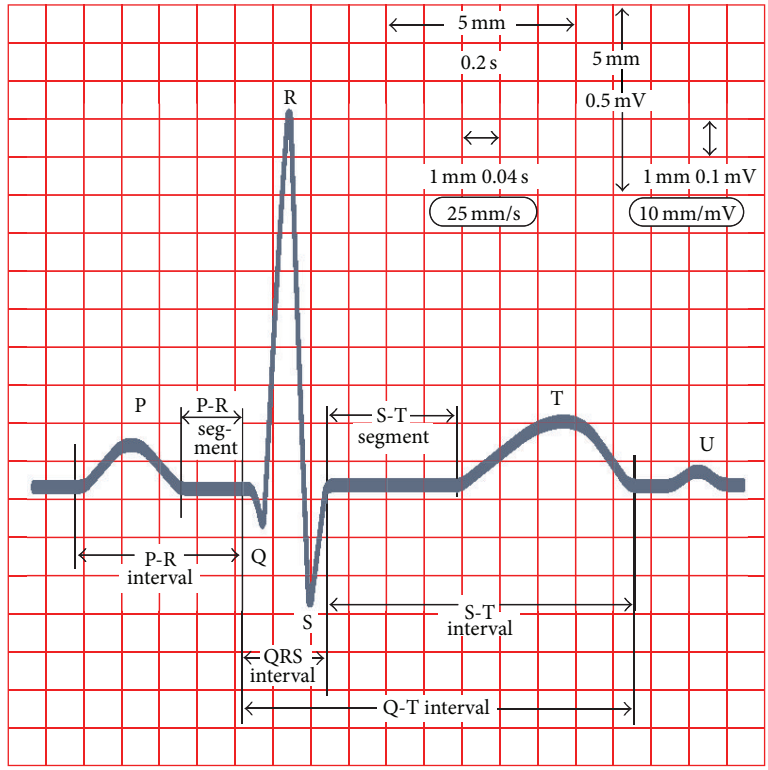


Figure 2: Stylized shape of a physiological single beat, recorded on ECG graph paper. Main relevant points, segments, and waves are highlighted.

when the ventricles recharge for the next contraction, named repolarization (see [12] for more details on ECG).

Concerning the ECG detection of AF events, characteristic findings are the absence of P waves with unorganized electrical activity in their place and irregular RR intervals due to irregular conduction of impulses to the ventricles. While the analysis of P wave is quite complicate, the study of RR intervals is simpler. Hence, it could be an effective way to investigate the presence of AF and to detect when a single event starts and ends. Several examples exist in the literature (see [13–16]), which are focused on the peculiar variance of RR intervals during the AF process, and this variance is much greater than the one during the physiological heartbeat.

Anyway, in many situations, an AF event does not follow a physiological time slot but comes after other types of arrythmia. At the same time, in many cases, the irregular heartbeat does not disappear when the event finishes. According to these problems, it may be possible to look at an irregular heartbeat even when the AF event itself has not already started or has already finished. So, a method based on detection of changes in the variance of the process can lead to inaccurate results and can fail as described previously. Hence, methods which are not based on the analysis of the process variance are needed, in order to identify suitable quantities to characterize the different phases, say "pre- AF," "AF," and "post- AF." To this aim, efforts are usually focused

on changepoint detection of the spectrum or of the mean of a time series (see [17–20] among others). In these cases, the tachogram is considered as a time series (see [21, 22]), the order of the model is fixed, that is, orders p, d, and q of the autoregressive (AR), integrated (I), and moving average (MA) component, respectively, are previously established, and the focus is on the evolution of the estimated model parameters.

In this work, we assume that the tachogram, during an AF event, is characterized by a specific process. Hence we propose a different approach: we describe the phases of AF by means of ARIMA models characterized by different orders p, d, and q. The main issue becomes then to point out proper statistical methods for detecting changes in the order of the model. To achieve this goal, we firstly carry out a simulation study to test the new statistical method we propose, then we analyze data of 8 patients affected by AF. In particular we have for each patient the tachogram from two hours before to two hours after an event of AF. Although there are a lot of readings about the change point detection of time series, there is a lack of literature if the approach we just mentioned is considered.

The paper is organized as follows. In Section 2, we introduce some elements of time series processes theory related to ARIMA models used for modeling the RR intervals time series. We present the statistical method developed for identifying the AF event (Section 2.3), based on the analysis of multiple test P-values with an improvement of the Bonferroni correction, and we test it in a simulation setting (Section 2.4), in order to assess the performance of the proposed method. Then in Section 3, we present the results obtained applying our method to real data (tachograms of patients affected by AF). Section 4 contains discussion and conclusions.

All the simulations and the analyses of real data have been carried out using R statistical software [23].

2. Materials and Methods

In this section, we introduce ARIMA models [24] as a tool for modeling the RR time series dynamic. Then, we present the statistical techniques developed for identifying onset and end of AF events. Moreover, a simulation study is carried out to test the performance of the new method we propose, and results of simulations are discussed.

2.1. Autoregressive Integrated Moving Average (ARIMA) Models. Many empirical time series have no constant mean. Even so, they exhibit a sort of homogeneity in the sense that a suitable affine transformation could have constant mean. Models which describe such homogeneous nonstationary behavior can be obtained by supposing some suitable differences of the process to be stationary. Referring to the framework and theory treated in [24], we focus on the properties of the important class of models for which the dth difference ($\nabla^d z_t = z_t - z_{t-d}$) is a stationary ARMA process. Then, let us consider the model

$$\phi(B)\nabla^d z_t = \theta(B)\, a_t, \tag{1}$$

where B is the backward shift operator,

$$\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j, \qquad \theta(B) = 1 - \sum_{h=1}^{q} \theta_h B^h, \tag{2}$$

with ϕ_j, $j = 1, \ldots, p$ and θ_h, $h = 1, \ldots, q$ suitable parameters to be estimated. Generally these estimates are performed through ML methods [24]. Process (1) is an Autoregressive Integrated Moving Average (ARIMA) process. If the autoregressive operator $\phi(B)$ in (1) is of order p and the moving average operator $\theta(B)$ is of order q, then (1) is an ARIMA(p, d, q) process.

2.2. Model Diagnostic Checking. Suppose to fit model (1) obtaining ML estimates $(\widehat{\phi}, \widehat{\theta})$ for the parameters of interest. We will refer to the quantities

$$\widehat{a}_t = \widehat{\theta}^{-1}(B)\, \widehat{\phi}(B)\, \nabla^d z_t \tag{3}$$

as the residuals. As the number of observations increases, $\widehat{a}_t$ becomes closer to the white noise a_t. Now suppose p, d, and q were chosen correctly and that we knew the true parameter values ϕ and θ. Then, the estimated autocorrelation $r_k(a)$ of the process a would be distributed approximately normally with zero mean (see [25]). Now, in practice, the parameters p, d, and q are unknown and only the estimates $(\widehat{\phi}, \widehat{\theta})$ are available for calculating $\widehat{a}$. Then, autocorrelation $r_k(\widehat{a})$ of $\widehat{a}$ can yield valuable evidence concerning the lack of fit. An interesting way to analyze the goodness of fit of the model is then to consider the $r_k(\widehat{a})$ taken as a whole. Let us suppose that we have the first K autocorrelations $r_k(\widehat{a})$ $(k = 1, 2, \ldots, K)$ from any ARIMA(p, d, q) process. Then, it is possible to show (see [26]) that, if the fitted model is appropriate, the statistic

$$Q = \widetilde{n}(\widetilde{n} + 2) \sum_{k=1}^{K} \frac{r_k^2(\widehat{a})}{(\widetilde{n} - k)} \tag{4}$$

is approximately distributed as $\chi^2(K - p - q)$, where $\widetilde{n} = n - d$, with n equal to the number of observations. Therefore, an approximate test of the hypothesis of model adequacy may be performed. The statistic Q is called Ljung-Box statistic.

2.3. A Method to Detect Structural Changes in Time Series. We now consider a phenomenon that evolves according to an ARIMA process. We wish to analyse a time series and to detect when such a phenomenon starts and/or ends. If this specific phenomenon is characterized by a higher (or lower) variability with respect to the current situation, then there is a huge number of methods effective in detecting these changes in variability. Examples are control charts (see [27]) and methods based on graphical analysis among others (see [15]). However, there are a lot of situations in which a phenomenon is not characterized by a modification of the variability, but by some changes in the process that generates the observations. In these cases, methods such those mentioned earlier are useless and other methodologies have to be considered. For example, in the literature, there is a huge quantity of methods

that deal with structural changes in time series concerning changes of the mean or of the parameters values of the ARIMA model (see [19, 20] and the references therein). Nevertheless, we may be interested in dealing with a different situation. For example, we may consider a problem where the presence or the absence of a phenomenon is characterized not in a change of the parameters values of the model, but in a modification of the process itself. We wish to present here a method for dealing with this kind of situations.

As we mentioned before, our main goal is to identify the beginning and the end of a specific phenomenon modeled by an ARIMA process. This means firstly to identify the model parameters of the phenomenon under study, that is, the values of d, p, and q. As we have previously presented, in the case of a stationary model, the autocorrelation and partial autocorrelation function will quickly approach zero. Knowing that the estimated autocorrelation function tends to follow the behavior of the theoretical autocorrelation function, failure of this estimated function approaching zero rapidly might logically suggest that we should treat the underlying stochastic process as nonstationary in z_t, but possibly as stationary in $\nabla^d z_t$. Once identifying one or more possible values for d, we move to the choice of p and q. This may be done considering the specific behaviors of the autocorrelation and partial autocorrelation functions and corresponding cut-off lags (see [24] for the details).

To identify the starting and ending times of the phenomenon of interest, we propose the following procedure. Consider the first N observations (with N much smaller then the number n of observations) and fit the identified model on this subsample. Then, the P-value of the Ljung-Box test (choosing a value for K) is recorded. These operations have to be repeated over the sub-sample from the second to the $N+1$th observation. Once reaching the last observation, the procedure ends producing a "time series" of P-values which may be used to detect the beginning and the end of the phenomenon of interest.

The purpose is to test the null hypothesis that the phenomenon is present against the alternative hypothesis that the phenomenon is absent. This may be formalised as follows:

$$H_0 : p = \overline{p} \wedge d = \overline{d} \wedge q = \overline{q} \text{ versus}$$
$$H_1 : p \neq \overline{p} \vee d \neq \overline{d} \vee q \neq \overline{q}, \tag{5}$$

where $\overline{p}$, $\overline{d}$, and $\overline{q}$ are the parameters indicating the order of the ARIMA process related to the phenomenon under study. In order to build the critical region for the test (5), the first P-values, say M, can be considered, and the rejection region can be constructed through a multiple test procedure, where the adjustment for multiplicity is based on the correction proposed by Simes. So doing, the approximate level of the test is equal to α (see [28] for the detailed work). The decisional criterion is the following. After the M P-values have been ordered from the minimum (say $p_{(1)}$) to the maximum (say $p_{(M)}$), the null hypothesis is rejected if for at least one j from 1 to M the following inequality is satisfied:

$$p_{(j)} \leq \frac{j\alpha}{M}. \tag{6}$$

It can be proved that this procedure provides an approximate level equal to α. Furthermore the test results are less conservative than a test implemented using a classical Bonferroni correction, especially in this situation, where single tests are highly correlated.

The method to detect start and/or end of a specific phenomenon follows these steps:

(1) implement the test in (5)-(6) over the first M P-values, and at the $N+M-1$th observation, the output is set to 0 if there is statistical evidence to reject the null hypothesis, while it is set at 1 otherwise;

(2) repeat step (1) after a shift of one observation until the last one is reached.

Once the procedure ends, an output of $0'$s and $1'$s is available. Also, 1 indicates the presence of the phenomenon, 0 the absence. Starting and end points can be then detected through this last 0/1 time series.

2.4. Simulation Study. In order to validate the proposed method, different situations have been tested and analysed. The main goals are the following:

(i) to point out settings where our method performs at best,

(ii) to assess the robustness of the method varying α and N,

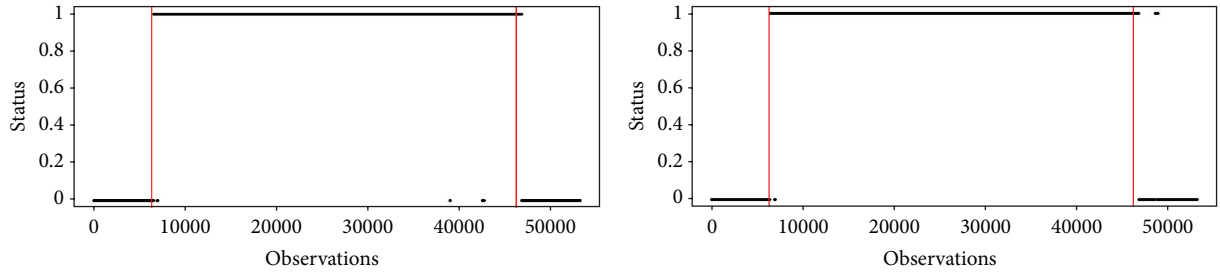
(iii) to make a sensitivity analysis over the parameter K of the Ljung-Box statistics.

The method presented in this paper is a technique to detect modification in the process underlying the observed phenomenon. We chose an ARIMA $(0,1,1)$ as Reference Process (RP), considering a sequence of 7000 realizations from a process, say P_{pre}, then 40000 realizations from the reference model, and finally 7000 realizations from another different process, say P_{post}. For all the simulations, the value of M was fixed equal to 100. In this particular case, test (5) becomes.

$$H_0 : p = 0 \wedge d = 1 \wedge q = 1 \text{ versus } H_1 : p \neq 0 \vee d \neq 1 \vee q \neq 1. \tag{7}$$

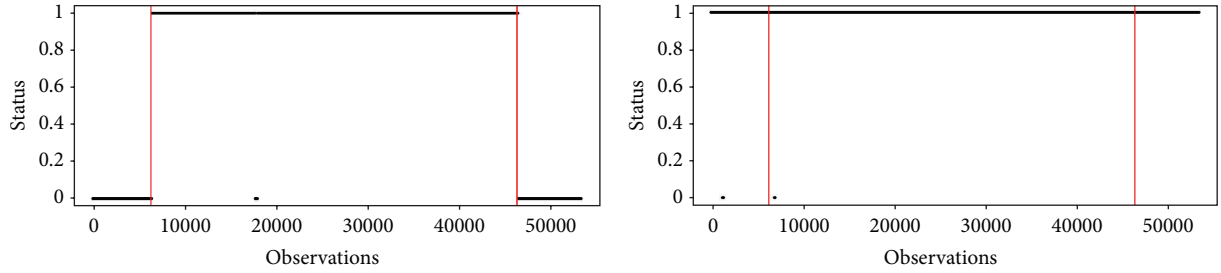
We tested it in different situations; in the first, second, and third simulations, P_{pre} and P_{post} are very different from RP, whereas in the fourth one they are quite similar. In particular we set $P_{\text{pre}} \equiv P_{\text{post}}$, and we considered an ARIMA $(4,1,2)$, ARIMA $(5,1,3)$, ARIMA $(2,2,0)$ and ARIMA $(1,1,1)$, respectively.

The parameters values for these simulations have been chosen randomly, under the constraint that the models were admissible. Their values are reported in Table 1. Figures 3(a), 3(b), and 3(c), obtained fixing $K = 5$, $N = 600$, and $\alpha = 0.01$, show that our method works very well in the first 3 settings, where the correspondence among the real starting and end points (red lines) and the 0/1 sequence is visible. In the fourth simulation, instead, the method is less able to catch the phenomenon under study, as it is shown in Figure 3(d).



(a) Output of the method: before and after the phenomenon under study, the process is an ARIMA (4,1,2)

(b) Output of the method: before and after the phenomenon under study, the process is an ARIMA (5,1,3)



(c) Output of the method: before and after the phenomenon under study, the process is an ARIMA (2,2,0)

(d) Output of the method: before and after the phenomenon under study, the process is an ARIMA (1,1,1)

Figure 3: Analysis of the output of the method changing the process underlying the observations before and after the phenomenon. Red lines represent the start and the end of the phenomenon.

Table 1: Parameters values used in the simulations. The first four models refer to P_{pre} and P_{post}, while the fifth model refers to RP.

ARIMA	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	θ_1	θ_2	θ_3
(4,1,2)	0.52	0.35	−0.04	0.11	/	−0.07	0.12	/
(5,1,3)	−0.66	−0.3	0.24	0.01	0.14	−0.08	−0.19	−0.29
(2,2,0)	−0.08	−0.25	/	/	/	/	/	/
(1,1,1)	−0.15	/	/	/	/	0.12	/	/
(0,1,1)	/	/	/	/	/	0.3	/	/

Table 2: Empirical type-I error probability varying N and the nominal value α.

	$N = 400$	$N = 600$	$N = 800$
$\alpha = 0.01$	0.004547	0.005300	0.004969
$\alpha = 0.05$	0.025889	0.028244	0.027377
$\alpha = 0.1$	0.051221	0.055458	0.058005

In the following, we focus on the case related to Figure 3(b), where the generating process is an ARIMA $(0, 1, 1)$, anticipated and followed by a process of observations generated from an ARIMA $(5, 1, 3)$. Cases (3a) and (3c) give similar results. We analyse how the power of the test in (7) is affected by α and N. For this analysis, we considered $K = 5$. If α was the real probability of the type-I error, the power would increase as α grows. We do not have the real probability of the type-I-error, but only an upper estimate. Nevertheless, we would observe the power growing as α increases. Another parameter that affects the power of the test is N. Again,

Table 3: Empirical power varying N and the nominal value α.

	$N = 400$	$N = 600$	$N = 800$
$\alpha = 0.01$	0.676305	0.873391	0.960046
$\alpha = 0.05$	0.828791	0.945367	0.987309
$\alpha = 0.1$	0.880587	0.967746	0.993346

the bigger N is, the greater the power of Ljung-Box test is. So, also the power of the global test should be raised. In Figure 4, the output of the method varying α (along the rows) and N (along the columns) is shown. It can be inferred that the behavior of the method is consistent, since the number of errors before and after the phenomenon decreases as α and N increase, as we expected.

We consider, for different values of α and N, the empirical type-I error probability and the empirical power computed over 40 simulations. Table 2 shows that this test is conservative, but the empirical type-I error probability is not so far from the nominal level of the test. Moreover, the results presented in Table 3 suggest that, once α is fixed, it is possible to increase the power of the test tuning N in a suitable way. Then, one could think to set a very high value of N in order to obtain a satisfactory power. However, this is not costless. In fact, the higher the value of N is the greater the delay in starting and ending points detection is.

Hence, the choice of the parameter N is regulated by a tradeoff between the desired power of the test and the delay in the detection of the phenomenon. To conclude the simulations' analysis, we would like to infer about

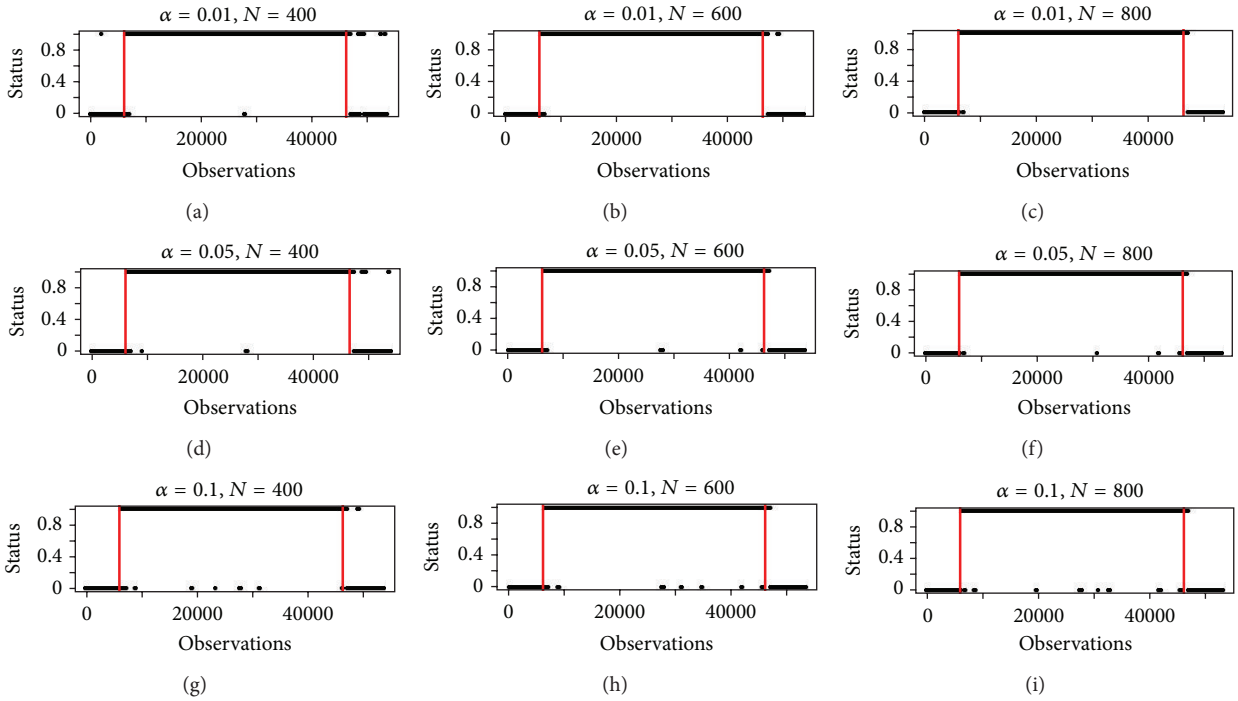


Figure 4: Output of the method varying α (along the rows) and N (along the columns).

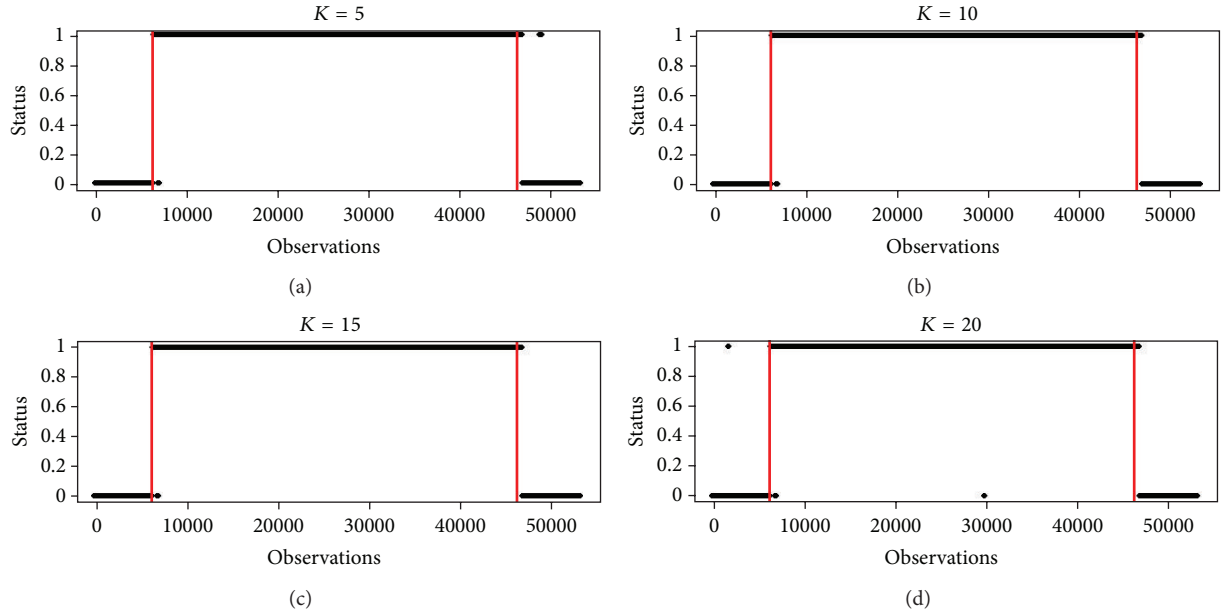


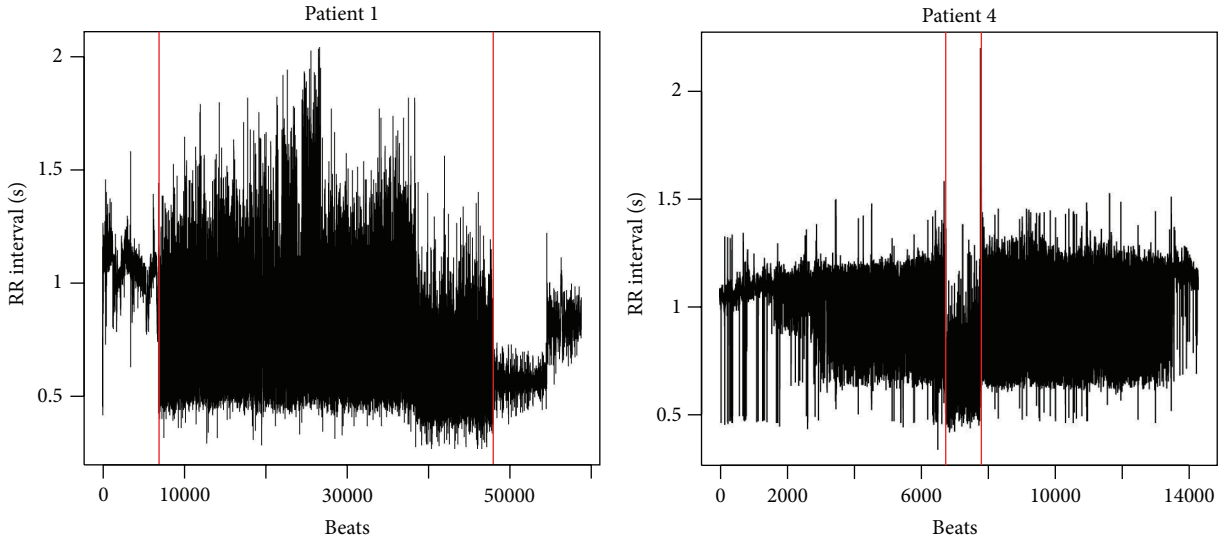
Figure 5: Output of the method varying K.

the parameter K of the Ljung-Box statistics in order to understand if the method is affected by a modification of its value. Let consider the situation where observations before and after the phenomenon were generated by an ARIMA $(5, 1, 3)$, and fix $\alpha = 0.01$ and $N = 600$. In Figure 5, the output of the method for different values of K (equal to 5, 10, 15, and 20, resp.) is shown. Although the outputs are different, no pattern of dependence on K appears.

3. Results and Discussion

Let us consider now an application of the method proposed in this paper to real data. Specifically we analysed RR intervals of 8 patients during Atrial Fibrillation (AF).

Data have been supplied to authors by Professor Luca Mainardi responsible of the Biomedical Signal Processing Laboratory of the Department of Bioengineering, Politecnico



(a) Tachogram of patient 1

(b) Tachogram of patient 4

Figure 6: Tachogram of two patients. For patient 1 (a), AF event comes after and is followed by normal sinus rhythm, characterized by low heart rate variability. Patient 4 (b) presents a high rate variability even before and after the AF event.

Table 4: Duration and number of beats of the event of AF.

Pat. no.	Duration AF (min.)	Beats
1	521	41085
2	613	43178
3	433	52937
4	13	1066
5	56	4326
6	442	52661
7	319	28229
8	229	17989

di Milano. Before patients underwent an ablation intervention, a seven-day Holter trace had been recorded using a one-channel *Del Mar Reynolds* Holter recorder, with sample frequency equal to 128 Hz. This protocol of data collection is in accordance with the Declaration of Helsinki for research with human beings. The data available are the RR intervals of such patients from two hours before to two hours after an event of AF. The duration of the phenomenon is different between patients and it is displayed in Table 4.

We want to detect the event of AF from the study of tachogram series. In some cases, the variability of RR intervals during AF is very high with respect to the physiological heartbeat. However, this remarkable change in the variability of the phenomenon could be absent, as highlighted in Figure 6. This is an example where the traditional methods based on detection of changes in the process variability are ineffective in detecting AF starting point.

The first step consists in the identification of a model for the RR intervals during AF. We used the autocorrelation and partial autocorrelation functions to determine a suitable model. As it is shown in Figure 7, the autocorrelation function of ∇z_t is truncated after the lag number one, while that of $\nabla^2 z_t$ is zero after the lag two. This behavior is typical of an ARIMA $(0, 1, 1)$ and $(0, 2, 2)$. Then, we set RP $\equiv$ ARIMA $(0, 1, 1)$. The same analysis done on the RR time series of pre- and post-AF does not lead to the same conclusions. Indeed autocorrelation and partial autocorrelation functions do not highlight these characteristics. Hence, the assumption that during Atrial Fibrillation the stochastic process generating the RR time series is different from the one that models other phases seems reasonable. Then, we would like to analyse the performance of the method in detecting start and end of such a phenomenon.

In order to achieve this goal, let us fix the following values for parameters: $K = 5$, $\alpha = 0.001$, and $M = 100$. Since in Section 2.4 the parameter N has been highlighted as the most important in affecting the performance of the proposed method, we analyse the output as N varies. In Figure 8, the outputs of the method applied to patients 1 and 5 are shown. We present here only the output for two patients, because the results for the other patients are quite similar.

Some considerations can be extrapolated observing Figure 8. First, we may point out to the behavior of the method as N increases. It can be seen that this behavior is in agreement with conclusions drawn from simulations. Then, we may analyse the delay in the detection of start and end of AF and the number of errors.

Dealing with the delay, since each observation is the time between an R peak and the following one, we can evaluate the time of the delay in the detection of the event of AF and not only the number of observations. As it is shown in Table 5, the delay in detecting the phenomenon is negligible if compared with the duration of AF, except for patient 4 affected by a very short AF event. Moreover, in some cases, the method is able to detect the AF event in advance.

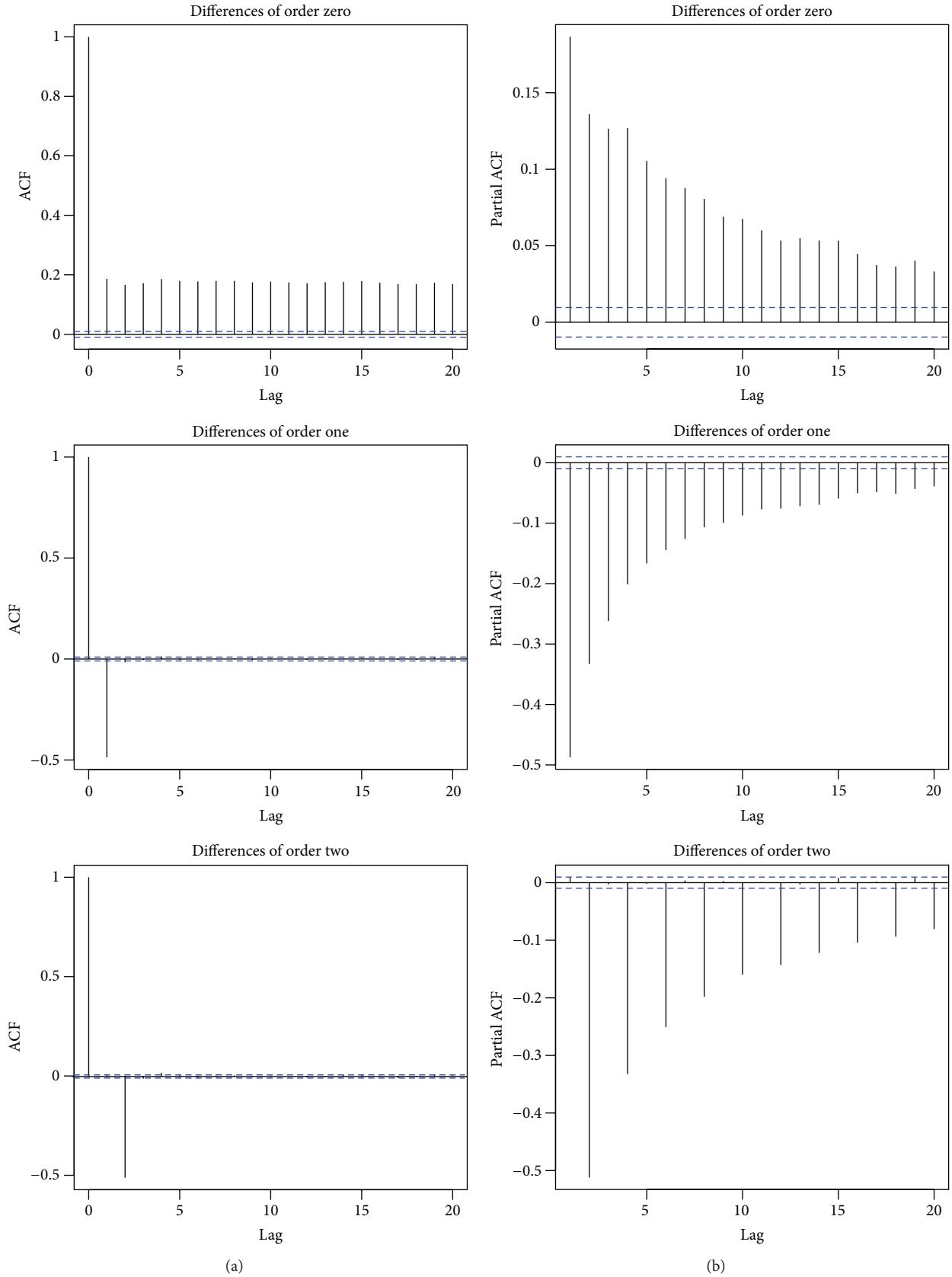


Figure 7: Patient 1: autocorrelation (a) and partial autocorrelation (b) functions for the time series of RR intervals, of the differences of order one and of the differences of order two.

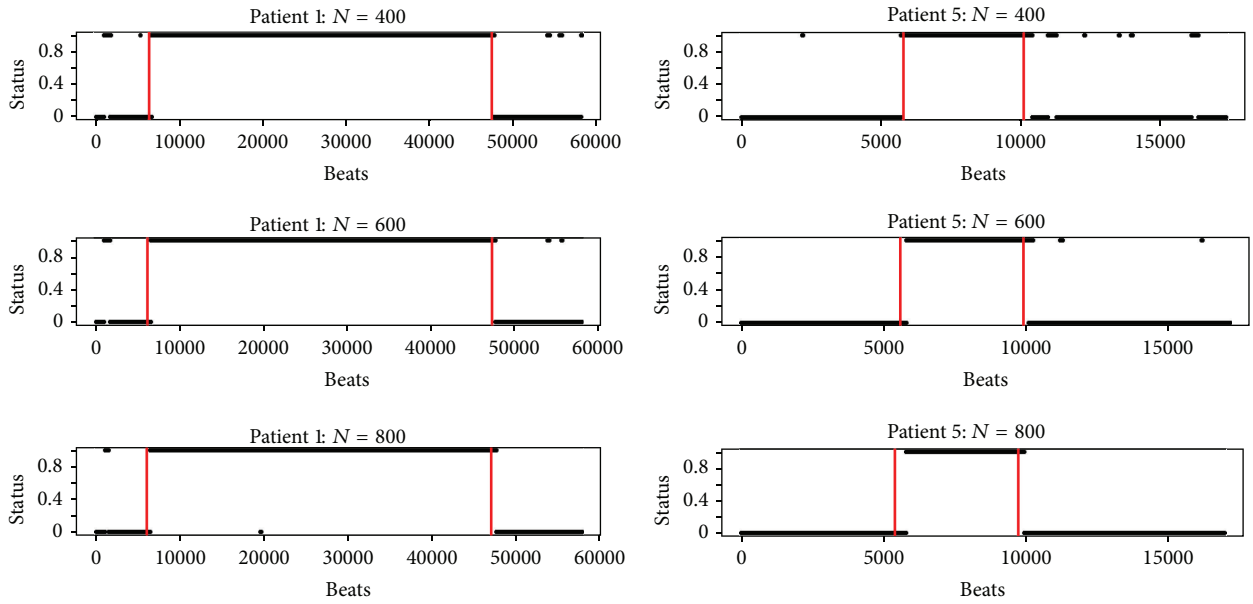


Figure 8: Output of the method for the patients 1 (a) and 5 (b) varying N.

Table 5: Delays of the method's output.

(a) Delays detecting the start of AF

Pat. num.	$N = 400$ (min.)	$N = 600$ (min.)	$N = 800$ (min.)
1	4.3	4.9	5.4
2	4.5	6.1	7.3
3	−2.4	0.2	−4.6
4	3.9	5.9	8.4
5	−1.4	2.7	5.6
6	−2.2	1	2.8
7	16.6	16.8	29.1
8	4.8	6.1	7.5

(b) Delays detecting the end of AF

Pat. num.	$N = 400$ (min.)	$N = 600$ (min.)	$N = 800$ (min.)
1	3.2	4.6	6
2	5.6	8.5	12
3	6.8	7.2	8.9
4	7.3	9.8	10.1
5	5.1	5	3.3
6	−3.3	−6.9	−6.3
7	3.2	5.3	7
8	3.3	5.3	7.2

Another important point we want to focus on is the number of errors made by the proposed method. From a first insight of Figure 8, we can observe that the most part of the errors seems to involve a few number of consecutive observations.

Table 6: Number of errors before (bef. corr.) and after (aft. corr.) the introduction of the artificial time delay (we fixed $N = 600$).

Pat. no.	Type-I errors (bef. corr.)	Type-I errors (aft. corr.)	Type-II errors (bef. corr.)	Type-II errors (after corr.)	Duration AF (min.)
1	0	0	4	1	521
2	1	0	0	0	613
3	16	6	1	1	433
4	0	0	4	4	13
5	1	0	2	0	56
6	23	5	3	1	442
7	8	1	8	3	319
8	0	0	10	6	229
Total	49	12	32	16	

Then, a correction can be implemented in order to reduce the number of errors (in this case, the whole time interval detected in a wrong way is considered as an error). We introduced an artificial time delay: after the first instant of output switching from zero to one (or vice versa), we wait for a given time to declare the AF event started (or ended); only if after this time the method is still indicating the presence (or absence) of the phenomenon, we can detect it. The introduction of this correction and its duration are problem driven. Since AF is not a dead risk pathology, the problem concerning the number of errors is more important than the detection delay, and so we chose to insert an artificial time delay of 3 minutes. Doing that, we decreased considerably the number of errors, as shown in Table 6.

4. Conclusions

In this paper, we proposed a statistical tool to identify starting and ending points of an event of AF (a common cardiac arrhythmia characterized by an irregular heartbeat) starting from the analysis of the RR intervals series. We presented a method based on time series analysis, and we performed a statistical test to automatically recognize the phases "pre-AF," "AF," and "post AF," especially in those situations where the AF event does not follow a physiological time slot and/or the irregular heartbeat is still present when the event finishes. The novelty of this work consists in looking at a structural change of the order (p, d, or q) of ARIMA model fitted on the RR time series for the AF event with respect to the "pre-AF" and "post-AF" phases. We tested the proposed method on different simulated data, taking a reference ARIMA model for the AF phase, and varying the model of "pre-AF" and "post-AF" phases.

Then, we applied the method to real RR intervals data. The results we obtained confirmed the goodness of the proposed method, which seems to be able to identify starting and ending points of an event of AF even when AF follows or comes before irregular heartbeat time slots. This is the innovative feature of our method, because the large variety of techniques that deal with the detection of AF do not take into account this particular situation. Since our method analyzes structural changes of the order of the ARIMA model, it can detect AF episodes also in those particular cases when before and/or after the AF event the heartbeat does not follow a normal sinus rhythm, characterized by a significative lower variability. This fact confirms that this methodology may become a helpful tool for the online and/or offline detection of AF. In particular this method could be useful in an offline control of Atrial Fibrillation events, such as a Holter monitor that is a prolonged type of ECG tracing. Since the traditional detection of AF through the analysis of the P wave might be long and hard and, in general, it is simpler to extract the RR intervals from a Holter, the proposed method could represent an automatic diagnostic tool that simplifies the detection of AF events.

Acknowledgments

The authors wish to thank Professor Luca Mainardi responsible of the Biomedical Signal Processing Laboratory of the Department of Bioengineering, Politecnico di Milano, for supplying the data and Dr. Valeria Vitelli for technical support in the statistical analysis.

References

[1] M. Malik and A. J. Camm, Eds., *Heart Rate Variability*, Futura Publishing Company, Armonk, NY, USA, 1995.

[2] M. Malik, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[3] N. Chattipakorn, T. Incharoen, N. Kanlop, and S. Chattipakorn, "Heart rate variability in myocardial infarction and heart failure," *International Journal of Cardiology*, vol. 120, no. 3, pp. 289–296, 2007.

[4] T. Kinoshita, T. Asai, T. Ishigaki, T. Suzuki, A. Kambara, and K. Matsubayashi, "Preoperative heart rate variability predicts atrial fibrillation after coronary bypass grafting," *The Annals of Thoracic Surgery*, vol. 91, no. 4, pp. 1176–1182, 2011.

[5] L. Mourot, N. Tordi, M. Bouhaddi, D. Teffaha, C. Monpere, and J. Regnard, "Heart rate variability to assess ventilatory thresholds: reliable in cardiac disease?" *European Journal of Preventive Cardiology*, vol. 19, pp. 1272–1280, 2012.

[6] B. Xhyheri, O. Manfrini, M. Mazzolini, C. Pizzi, and R. Bugiardini, "Heart rate variability today," *Progress in Cardiovascular Diseases*, vol. 55, pp. 321–331, 2012.

[7] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Berger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *Science*, vol. 213, no. 4504, pp. 220–222, 1981.

[8] G. Baselli, D. Bolis, S. Cerutti, and C. Freschi, "Autoregressive modeling and power spectral estimate of R-R interval time series in arrhythmic patients," *Computers and Biomedical Research*, vol. 18, no. 6, pp. 510–530, 1985.

[9] R. G. Mark and G. B. Moody, "ECG arrhythmia analysis: design and evaluation strategies," in *Advances in Processing and Pattern Analysis of Biological Signals*, I. Gath and G. F. Inbar, Eds., chapter 18, pp. 251–272, Plenum Press, New York, NY, USA, 1996.

[10] S. S. Chugh, J. L. Blackshear, W. K. Shen, S. C. Hammill, and B. J. Gersh, "Epidemiology and natural history of atrial fibrillation: clinical implications," *Journal of the American College of Cardiology*, vol. 37, no. 2, pp. 371–378, 2001.

[11] A. Houghton and D. Gray, *Making Sense of the ECG*, 1997.

[12] A. E. Lindsay, "ECG learning centre," 2006, http://ecg.utah.edu/.

[13] B. K. Bootsma, A. J. Hoelsen, J. Strackee, and F. L. Meijler, "Analysis of R-R intervals in patients with atrial fibrillation at rest and during exercise," *Circulation*, vol. 41, no. 5, pp. 783–794, 1970.

[14] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, and J. Habetha, "Detection of Atrial Fibrillation using model-based ECG analysis," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, December 2008.

[15] L. Mainardi, L. Sornmo, and S. Cerutti, *Understanding Atrial Fibrillation: the Signal Processing Contribution*, Morgan and Claypoll Publishers, 2008.

[16] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals," *Medical and Biological Engineering and Computing*, vol. 39, no. 6, pp. 664–671, 2001.

[17] A. Boardman, F. S. Schlindwein, A. P. Rocha, and A. Leite, "A study on the optimum order of autoregressive models for heart rate variability," *Physiological Measurement*, vol. 23, no. 2, pp. 325–336, 2002.

[18] R. A. Davis, D. Huang, and Y. C. You, "Testing for a change in the parameter values and order in autoregressive models," *The Annals of Statistics*, vol. 23, no. 1, pp. 282–304, 1995.

[19] S. Ling, "Testing for change points in time series models and limiting theorems for NED sequences," *The Annals of Statistics*, vol. 35, no. 3, pp. 1213–1237, 2007.

[20] D. Picard, "Testing and estimating change-points in time series," *Advances in Applied Probability,*, vol. 17, no. 4, pp. 841–867, 1985.

[21] S. Cerutti, L. T. Mainardi, A. Porta, and A. M. Bianchi, "Analysis of the dynamics of RR interval series for the detection of Atrial Fibrillation episodes," in *Proceedings of the 24th Annual Meeting on Computers in Cardiology*, pp. 77–80, September 1997.

[22] C. Cammarota and E. Rogora, "Independence and symbolic independence of nonstationary heartbeat series during atrial fibrillation," *Physica A*, vol. 353, no. 1–4, pp. 323–335, 2005.

[23] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, http://www.r-project.org/.

[24] P. Diggle, *Time Series: A Biostatistical Introduction*, Clarendon Press, Oxford, UK, 1990.

[25] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American Statistical Association*, vol. 65, pp. 1509–1526, 1970.

[26] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[27] D. C. Montgomery, *Introduction to Statistical Quality Control*, McGraw-Hill, New York, NY, USA, 2005.

[28] R. J. Simes, "An improved bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 73, no. 3, pp. 751–754, 1986.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 617604, 12 pages
http://dx.doi.org/10.1155/2013/617604

Research Article

Normality Index of Ventricular Contraction Based on a Statistical Model from FADS

Luis Jiménez-Ángeles,[1] Raquel Valdés-Cristerna,[2] Enrique Vallejo,[1] David Bialostozky,[1] and Verónica Medina-Bañuelos[2]

[1] *Nuclear Cardiology Department, Instituto Nacional de Cardiología "Ignacio Chávez", Juan Badiano No. 1 Colonia Seccion XVI, Tlalpan, 14080 Mexico City, DF, Mexico*

[2] *Neuroimaging Laboratory, Electrical Engineering Department, Universidad Autónoma Metropolitana-Iztapalapa, San Rafael Atlixco No. 186 Colonia Vicentina, Iztapalapa, 09340 Mexico City, DF, Mexico*

Correspondence should be addressed to Luis Jiménez-Ángeles; luis.jimenez@ieee.org

Received 27 September 2012; Accepted 22 February 2013

Academic Editor: Angel García-Crespo

Radionuclide-based imaging is an alternative to evaluate ventricular function and synchrony and may be used as a tool for the identification of patients that could benefit from cardiac resynchronization therapy (CRT). In a previous work, we used Factor Analysis of Dynamic Structures (FADS) to analyze the contribution and spatial distribution of the 3 most significant factors (3-MSF) present in a dynamic series of equilibrium radionuclide angiography images. In this work, a probability density function model of the 3-MSF extracted from FADS for a control group is presented; also an index, based on the likelihood between the control group's contraction model and a sample of normal subjects is proposed. This normality index was compared with those computed for two cardiopathic populations, satisfying the clinical criteria to be considered as candidates for a CRT. The proposed normality index provides a measure, consistent with the phase analysis currently used in clinical environment, sensitive enough to show contraction differences between normal and abnormal groups, which suggests that it can be related to the degree of severity in the ventricular contraction dyssynchrony, and therefore shows promise as a follow-up procedure for patients under CRT.

1. Introduction

Heart failure (HF) is defined as a complex clinical syndrome that can result from any structural or functional cardiac disorder and that impairs the ability of the ventricle to fill or eject blood [1]. According to a 44-year followup of the National Heart, Lung, and Blood Institute's Framingham Heart Study, approximately 5.7 million patients have an HF diagnosis in the United States. After HF is diagnosed the survival rate is lower in men than in women, less than 15 percent of women survive more than 8–12 years and the one-year mortality rate reaches 20% [2].

Ventricular dyssynchrony has also been associated with increased mortality in HF patients [3, 4]. Dyssynchronous contraction can be palliated by electrically activating in a synchronized form the right and left ventricles with a multisite pacemaker device. This kind of treatment is called cardiac resynchronization therapy (CRT). Several clinical studies have shown that CRT contributes to an increase in the life expectancy of subjects diagnosed with cardiac failure, specifically of the type where the left ventricle ejection fraction is under 35% or classified in levels III or IV, according to the New York Heart Association [5–7] criteria. In a meta-analysis of several CRT trials, evidence showed that HF hospitalizations were reduced by 32% and that all-cause mortality decreased by 25% after approximately 3 months of therapy [8]. In a randomized controlled trial comparing optimal medical therapy alone with optimal medical therapy plus CRT (without a defibrillator), CRT significantly reduced the combined risk of death by any cause and decreased the unplanned hospital admission for a major cardiovascular event by 37% [9].

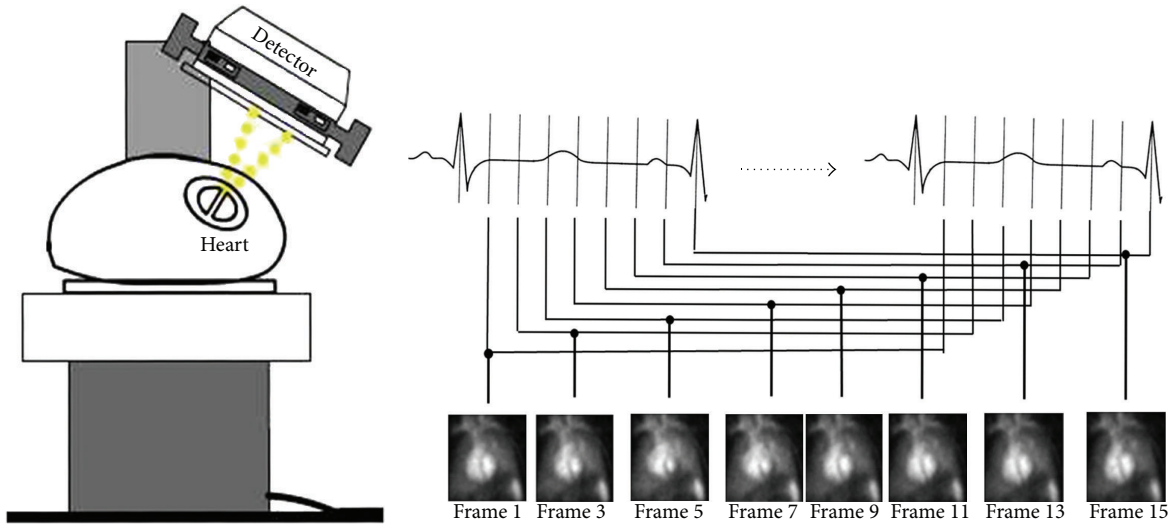


Figure 1: Schematic ERNA image acquisition. (a) Detector in the left anterior oblique (LAO) position to visualize the best RV and LV definition. (b) Several EKG-gated temporal frames corresponding to different phases of cardiac cycle are acquired in the LAO position. The ERNA images are obtained from summation of individual frames.

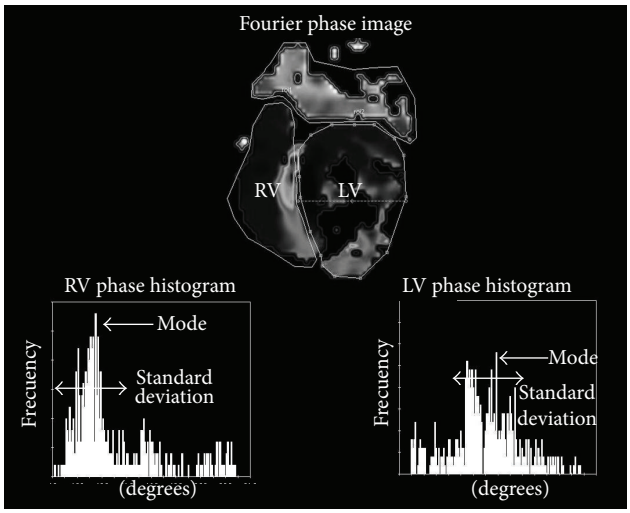


Figure 2: Parametric Fourier phase image and the phase image histogram for both ventricles (RV and LV). Indices like mean, standard deviation, and mode, computed from the statistical distribution of the phase angles, have been proposed to evaluate contraction abnormality patterns.

However, 20% to 30% of patients having HF do not benefit from resynchronization therapy, probably due to several causes [10].

(i) The established criteria to select candidates for CRT are limited.

(ii) The stimulation leads are not properly placed.

(iii) There is excessive fibrous tissue at the stimulus location.

This has led to the definition of several indexes, extracted from imaging modalities, to measure the cavities mechanical contraction, and that allow the proper identification of candidates to undergo CRT [11]. The quantification of peak systolic velocity and myocardial deformation from echocardiographic images has been proposed as representative indexes to evaluate ventricular dyssynchrony. However, studies carried out in multiple health centers have shown that it has a very low sensitivity to discriminate subjects who respond to CRT from those who do not [12]. Recent studies have reported indices, extracted from MRI, which can contribute to the solution of this problem, but the use of this imaging modality is restricted, depending on the type of resynchronization device that has been implanted [11, 13–16].

Radionuclide-based imaging is another alternative to evaluate ventricular contraction synchrony [17]. The equilibrium radionuclide angiography (ERNA) is a set of images that represent the spatial distribution of a radiotracer and relates pixel's intensity to ventricular volume. The general setup for ERNA image acquisition (see Figure 1) consists in locating the detector of the gamma camera in the left anterior oblique view (LAO) with patient at rest in supine position after the injection of red blood cells marked with Tc-99m. Synchronized acquisition of images with the electrocardiogram (EKG) allows the accumulation of radioactivity in several R-R intervals, to construct the image set that represents a specific instant of the cardiac cycle [18, 19].

ERNA images are processed by adjusting the first harmonic component of the Fourier Transform (FT) of each pixel's temporal intensity evolution (time-activity curve (TAC)). From these components, phase angles, which are representative of the TAC behavior, are extracted and a map (phase image) of the ventricular contraction sequence is constructed [20]. Several indices, taken from the statistical distribution of the phase angles, have been proposed to detect contraction abnormality patterns [21–23]. Figure 2 shows an example of the phase image corresponding to

an abnormal contraction pattern (with intraventricular and interventricular dyssynchrony), together with the right (RV) and left (LV) ventricle histograms. The modes and standard deviations of these distributions are measured and used as clinical indices to identify abnormalities.

The standard deviation of the pixels' phase angles measured in each ventricular Region of Interest (ROI) represents intraventricular dyssynchrony and the difference between the means of the phase angles of both ventricular ROI represents interventricular dyssynchrony [24]. Several studies have reported an improvement of interventricular and intraventricular dyssynchrony after CRT, using the Fourier phase analysis of ERNA images [25–29]. Dauphin et al. [30] showed that interventricular dyssynchrony was identified as an independent predictive factor of good clinical response with a practical cut-off value of 25.5°, a sensitivity of 91.4%, and a specificity of 84.4%. However, Fourier phase analysis based only on one FT harmonic has its limitations, since it assumes periodic TACs and a smooth transition between the first and last frame of the dynamic images series. These drawbacks are more prominent in the regions with severe contraction pattern abnormalities.

Factor Analysis of Dynamic Structures (FADS) has also been proposed as a valuable tool to detect abnormalities in ventricular cavities' movement [31, 32]. It is applied to ERNA images to extract those TACs associated to the physiological behavior of a specific region and assumes that there are pixel clusters with the same temporal evolution which define their morphology. Therefore, FADS determines the TACs (coefficients) of pixel groups with the same behavior, in addition to their geometry and spatial location (factors) [33, 34]. In a previous work carried out by our group, we analyzed the contribution and spatial distribution of the most significant factors present in a dynamic series of ERNA images and we proposed an alternative method to reconstruct the phase image. In [35], we reported that more than 90% of the information contained in an image series is represented by the three most significant factors (3-MSF) and that the third factor increases considerably whenever an abnormality of the contraction pattern is observed. Also, a detailed analysis of the scatter plots of the 3-MSF showed the importance of the third factor to adequately separate regions having an abnormal contraction pattern. Therefore, the need to propose an index to quantify contraction abnormality, using the representative information extracted from dynamic image series, becomes evident.

In this work, a probability density function model of the 3-MSF, extracted from FADS for a control group, is presented; also a reference normality index, based on the likelihood between the control group's contraction model and a sample of normal subjects, is proposed. The index was then statistically compared with those computed for two populations of patients satisfying the clinical criteria to be considered as candidates for a CRT: a group with complete left bundle branch block (LBBB) and a group with dilated cardiomyopathy (DCM).

The paper is structured as follows: in the Methodology section we describe the proposed model to characterize a normal contraction pattern (Sections 2.1 and 2.2); the defined index to quantify the degree of normality with respect to a reference population (Section 2.3); the populations considered to test the proposed index (Sections 2.4 and 2.5) as well as the statistics employed (Section 2.6). The Results section describes the findings of the proposed normality index tested in different cardiopathies and compared to the clinical standard provided by Fourier phase analysis. These are analyzed and interpreted in the Discussion section, to finally conclude at the corresponding section.

2. Materials and Methods

2.1. Factor Analysis of Dynamic Structures (FADS). Let $\mathbf{X}_{\mathrm{TAC}}(p,q) = \mathbf{X}[(i,j),k]$ be a bidimensional array (Figure 3(c)), whose indices represent the (i,j)th pixel value of the kth frame of the acquired image series. Each frame size is $M \times M$ pixels ($p = (i-1) \times M + j, q = k$) as shown in Figure 3(a). $\mathbf{X}_{\mathrm{TAC}}(p,q)$ represents the time-series generated for each pixel on the image set, known as time-activity curves (TAC) (Figure 3(b)).

Let $\mathbf{Q}$ be a linear transformation that decorrelates the ERNA image set ($\mathbf{X}_{\mathrm{TAC}}$), so that

$$\begin{aligned} \mathbf{F} &= \mathbf{X}_m \mathbf{Q}, \\ \mathbf{Q} &= \mathbf{V}\mathbf{D}, \end{aligned} \tag{1}$$

where $\mathbf{F}$ are the factors of $\mathbf{X}_{\mathrm{TAC}}(p,q)$ (Figure 3(c)), $\mathbf{X}_m$ is $\mathbf{X}_{\mathrm{TAC}}(p,q)$ with the mean-value removed; $\mathbf{V}$ is the eigenvector set of the autocorrelation matrix of $\mathbf{X}_m$ and $\mathbf{D}$ is the scaled diagonal matrix of the eigenvalues set of the autocorrelation matrix of $\mathbf{X}_m$. The contribution of each factor is determined by the corresponding eigenvalue magnitude.

2.2. Normal Contraction Pattern Model. The three most significant factors (3-MSF) of the ERNA studies, obtained for a population of normal subjects, were analyzed; the spatial distribution of those factors can be observed in Figure 4. Every point in the factorial 3D space is associated to the projection of a given pixel in the ERNA image, into each of the main eigenvectors $\mathbf{V}$.

The probability density function (PDF) of those factors was modeled by a linear combination of R Gaussian density functions (Gaussian mixture), defined by the following expression [36]:

$$p(f) = \sum_{r=1}^{R} w_r N(f \mid \mu_r, \Sigma_r), \tag{2}$$

where f is observed variable, w_r is relative weight of the rth gaussian function of the mixture, and $N(\mu_r, \Sigma_r)$ is the multivariate Gaussian PDF with μ_r and Σ_r parameters

$$\sum_{r=1}^{R} w_r = 1, \quad 0 \le w_r \le 1. \tag{3}$$

To estimate the mixture parameters $\{w_r,\ \mu_r\ y\ \Sigma_r\}$ for $r = 1, \ldots, R$, the expectation maximization algorithm, that maximizes the mixture model likelihood, was used [37].

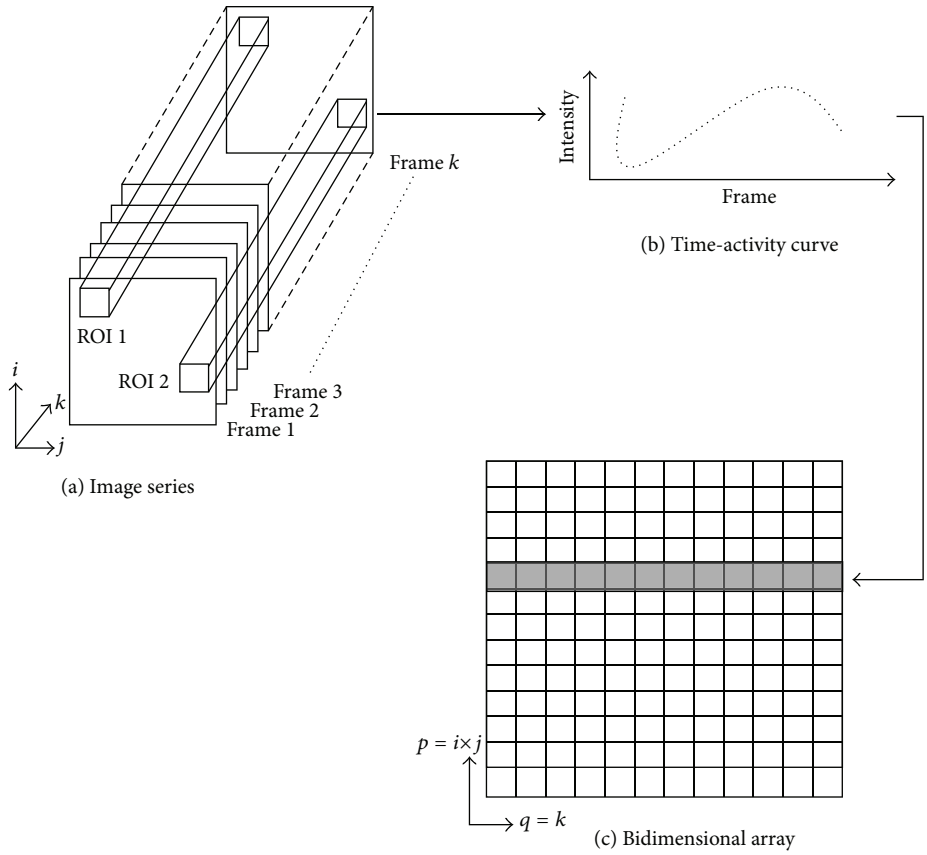


Figure 3: (a) ERNA study consisting of a k-images series, with frames having $i \times j$ pixels. (b) Time-activity curve extracted from a particular Region of Interest (ROI 1). (c) Bi-dimensional array constructed from the image series.

Six data groups were assembled, with 10 subjects each, randomly selected from a total of 23 control subjects, following a standard bootstrap resampling procedure. The PDF of the 3-MSF was modeled for each group, using the procedure described above. The number of components of the Gaussian mixture was determined considering the bayes information criterion (BIC) [38, 39]. The models' parameters, BIC, and likelihood were calculated using the R package (R Foundation, http://www.r-project.org/) [40].

2.3. Normality Index. Considering that the likelihood estimated on a data sample (f_s) represents the probability that those observations are well described by the assumed model (Gaussian mixture with w, μ, and Σ parameters), in this work we propose a normality index based on this probability. Assuming statistical independence between observations, the average log-likelihood of a sample set with respect to the reference (healthy) population can be defined as a comparative index (I_N) of a normal contraction pattern as follows:

$$I_N = \frac{1}{\|S\|} \sum_{s \in S} \log \left(\sum_{r=1}^{R} w_r N \left(f_s \mid \mu_r, \Sigma_r \right) \right), \quad (4)$$

where S is the observations' set, that in our case corresponds to the ventricular region TACs for each subject.

The normality index (I_N) for a group of eight normal subjects (not considered for the mixture parameters estimation) was measured and statistically compared with the indices obtained for LBBB and DCM subjects.

2.4. Studied Populations. Three subject groups were considered in this study: 15 subjects with LBBB; 13 patients with DCM and 31 normal subjects (23 as a control population for the training stage and 8 to define the normality index); all individuals gave their informed consent to participate in the study. The specific characteristics for these populations are shown in Table 1.

LBBB occurs whenever the electric impulse traveling from auricles to ventricles is interrupted, thus causing a QRS complex duration longer than 0.12 s. This delay provokes interventricular contraction asynchrony which can progress to eventually become cardiac insufficiency [41]. The LBBB studied population consisted of 15 asymptomatic subjects (8 males, 7 females), having a left ventricle ejection fraction (EF) greater than 45%, as determined by ERNA according to the New York Heart Association (NYHA) criteria [42]; the subjects did not present cardiovascular symptoms and did not have a previous history of myocardial infarct and/or cardiac insufficiency.

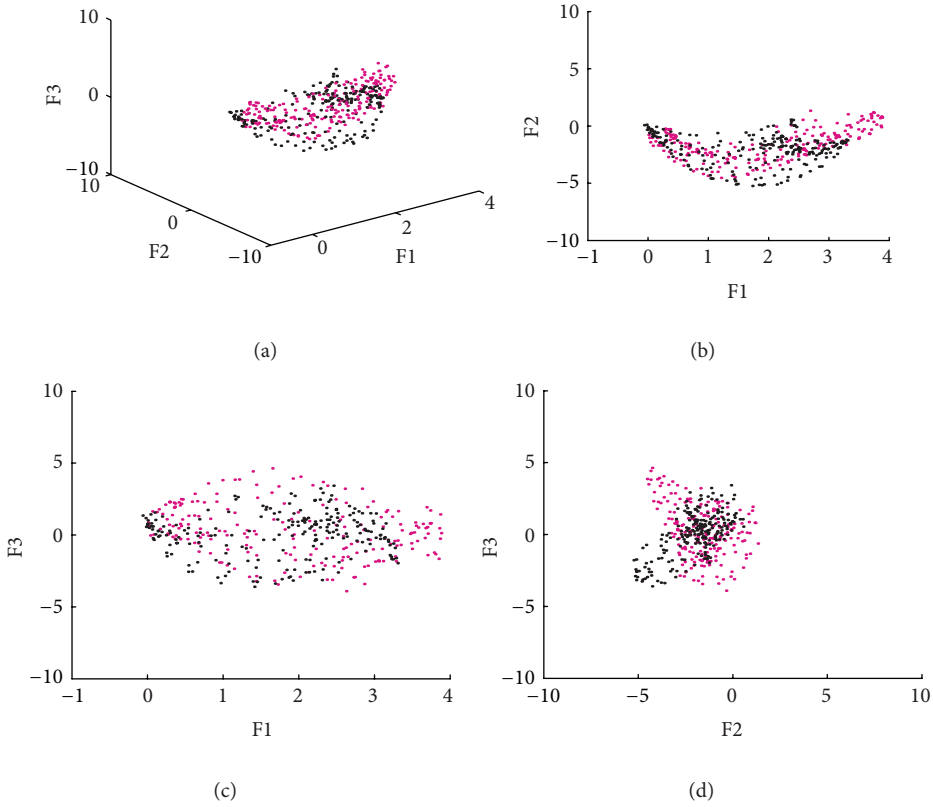


Figure 4: Scatter plot of the 3-MSF (F1, F2, and F3) calculated for a subject with a normal contraction pattern (right ventricle in magenta, left ventricle in black).

Table 1: Characteristics of the studied populations.

	Control ($n = 23$)	LBBB ($n = 15$)	DCM ($n = 13$)
Age (years)	28 ± 5	59.90 ± 9.09	45.6 ± 16.5
LVEF	60 ± 5.84	59.5 ± 9.4	22.2 ± 6.7
SAH *n* (%)	0	11 (73.3)	3 (20)
Diabetes mellitus *n* (%)	0	1 (6.6)	2 (13.3)
Dyslipidemia *n* (%)	0	3 (20)	4 (26.6)
Smokers *n* (%)	0	2 (13.3)	7 (46.7)

LVEF: left ventricle ejection fraction, SAH: systemic arterial hypertension, LBBB: left bundle branch block, DCM: idiopathic dilated cardiomyopathy.

Subjects with idiopathic DCM and cardiac insufficiency present left ventricle (LV) or right ventricle (RV) dilatation of unknown causes, inter- and intraventricular abnormal contractility; they must reunite all of the criteria to be considered as CRT candidates [41–44]. The DCM population consisted of 13 subjects, with EF of 22.2 ± 6.7%, as determined by ERNA, and with an average QRS duration of 0.160 ± 0.26 s; they also presented a class III or IV cardiac insufficiency, according to the NYHA [42].

The control population consisted of 23 volunteers (18 males, 5 females) having an EF of 60 ± 5.84%; with a low probability of coronary arterial disease and without a history of myocardial acute infarct. This group presented an EKG without abnormality and their cardiac function was considered normal, after a thorough clinical evaluation.

2.5. ERNA Images Acquisition. The same *General Electric* millenium MPR/MPS gamma camera was used for all ERNA image acquisition. It has a single head with 64 photomultiplier tubes and it is equipped with a low energy, high resolution parallel-hole collimator; the calibration of the energy peak was centered at 140 KeV and the detector's uniformity was guaranteed at less than 5% [45]. Images were digitized at a 64 × 64 pixels resolution and 1.33 zoom factors.

Erythrocytes were marked applying an *in vivo/in vitro* modified technique with 740 to 925 MBq of Tc-99m, using an UltraTag Kit [46, 47]. EKG was continuously monitored to synchronize images acquisition with the R wave. To eliminate ventricular extrasystoles during acquisition, a beat acceptance window was defined at ±20% of the average heart rate. Images were taken in an anterior left oblique projection, in order to simultaneously attain the best definition of left and right ventricles. A total of 16 frames were obtained with a density of 300 Kcounts per frame.

For each subject, an image corresponding to the end of diastole was selected and manually segmented by an expert, to define the ventricular area. This segmentation defines a mask that is used to automatically extract the ventricular regions from the other frames.

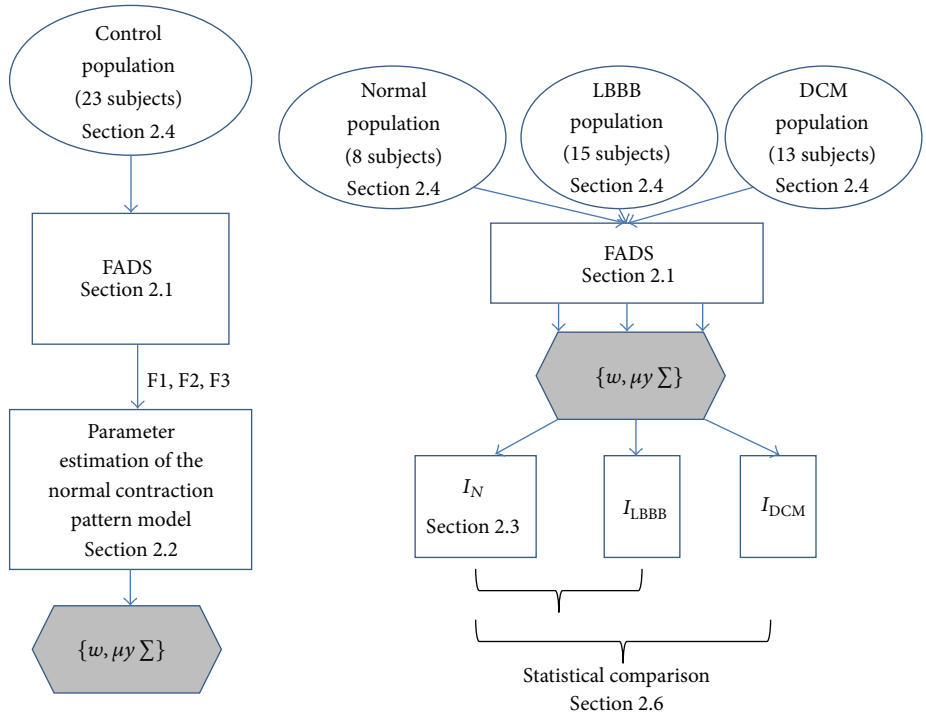


Figure 5: Schematic representation of the proposed methodology. (a). The normal contraction pattern was modeled as a linear combination of Gaussian density functions. The relative weights (w), means (μ) and covariance matrices (Σ) of the Gaussian mixture were estimated maximizing the mixture model likelihood of the probability density functions of the 3 most significant factors (F1, F2, and F3) computed from the Factor Analysis of Dynamic Structures (FADS) and following a standard bootstrap resampling procedure with a set of 23 control subjects. (b) Considering that the likelihood estimated on a data sample represents the probability that those observations are well described by the assumed model with w, μ, and Σ parameters, we propose indices based on this probability. A normality index (I_N) for a group of normal subjects (not considered for the mixture parameters estimation) was measured, and statistically compared with the indices obtained for LBBB and DCM populations.

2.6. Statistical Analysis. The normality indices are expressed as the mean value ± standard deviation. Indices measured for the normal group were independently compared to those obtained for LBBB and DCM populations, using a t-test for independent samples and considering $P \leq 0.01$. The SPSS version 10.0 software was used for all statistical analyses.

2.7. Summary. To summarize, the methodology is divided in two stages: training to obtain the model's parameters (Figure 5(a)) and application of this model to populations' comparison (Figure 5(b)).

3. Results

3.1. Factor Analysis. The information obtained for the 3-MSF (F1, F2, and F3) of the left and right ventricular regions was projected into scatter plots to observe differences between populations and between ventricular regions. Figures 4, 6, and 7 correspond to the Control, LBBB, and DCM populations, respectively.

The scatter plots obtained for the subjects studied show that the information for the left and right ventricles is overlapped in the control population (Figure 4), but also in the F1 versus F2 projection for abnormal contraction patterns (Figures 6(b) and 7(b)). However, in the presence of interventricular asynchrony, as in the case of the LBBB population, it was necessary to incorporate the F3 factor information in the analysis, to appreciate a clear separation between ventricular regions (Figures 6(a), 6(c), and 6(d)). Also, for the DCM population, that presents inter- and intraventricular asynchrony, the scatter plots that incorporate the third factor information (F3) show this left and right regions partition, although it was less evident than in the case of LBBB subjects, probably explained by the difference in the asynchrony type.

3.2. Model of the Factors' Probability Density Function. Different models were obtained for the PDF of the 3-MSF for six groups, with 10 subjects each, randomly selected from the control population. Table 2 shows the characteristics of the mixture of Gaussian functions that best adjusted each group.

For the control population, the minimum BIC and log-likelihood values corresponded to model 5, so that it has the highest probability of best describing the data corresponding to the PDF of the F1, F2, F3 distribution of the normal contraction pattern group. This model is characterized by

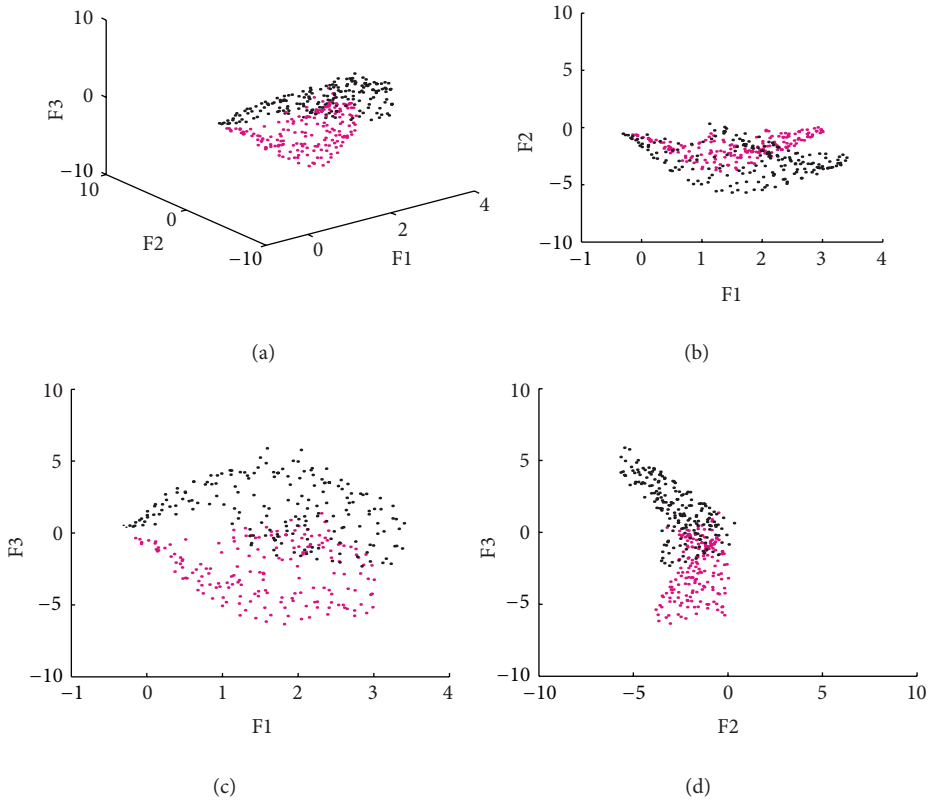


Figure 6: Scatter plot of the 3-MSF (F1, F2, and F3) calculated for a subject with a LBBB (right ventricle in magenta, left ventricle in black).

Table 2: Characteristics of the models that best describe the contraction pattern for the six defined groups.

Model	Type	Number of components	BIC	Log-likelihood
1	VVV	5	34258.15	16928.97
2	VVV	5	34094.83	16847.28
3	VEV	8	34502.55	16984.84
4	VEV	6	33676.04	16597.08
5	VVV	5	32863.10	16232.13
6	VEV	8	34502.55	16984.84

BIC: bayes Information criterion, Log-likelihood: logarithm of the likelihood value, VEV: Ellipsoidal, same shape, variable orientation; VVV: Ellipsoidal, varying volume, shape, and orientation.

having five Gaussian functions with variable volume, shape, and orientation. The weight parameters (w), mean values (μ), and covariance matrices (Σ) that describe the selected model are shown in Table 3.

The level curves of the adjusted model were superimposed with the information of the 3-MSF for one subject of each studied population. Figures 8, 9, and 10 show the correspondence for the normal, LBBB, and DCM subjects, respectively. The agreement between the model and the cardiopathic subjects is poor, as may be expected.

In Table 4, the normality indices (I_N) obtained for the populations studied, compared with the defined normal contraction model (see (4)), are presented. It can be observed that whenever the likelihood value increases and becomes statistically different from that calculated for the normal subjects, the probability that the model explains the data decreases. The calculated indices for the pathologic populations are statistically different from the I_N of the normal population, which suggests that the LBBB and DCM populations present abnormalities in the ventricular contraction pattern. Additionally, the DCM population presents a larger difference compared to the reference group; this was also corroborated with the clinical characteristics of the evaluated subjects and with the deterioration of their ventricular contraction pattern.

For comparison purposes, in Table 4 the most clinically used (mean, standard deviation, and mode) indices, extracted from phase analysis, are included. It can be observed that the standard deviation obtained from the traditional analysis also shows statistical differences between the normal and pathologic populations.

4. Discussion

Inter- and intraventricular contraction synchrony plays an important role in the heart pump function. The deterioration of contraction homogeneity can lead to a poor prognosis of clinical evolution, while a restoration of the ventricular contraction has proven to be of clinical benefit in patients with heart failure. However, despite the fact that several studies show that CRT can be of great benefit for severe

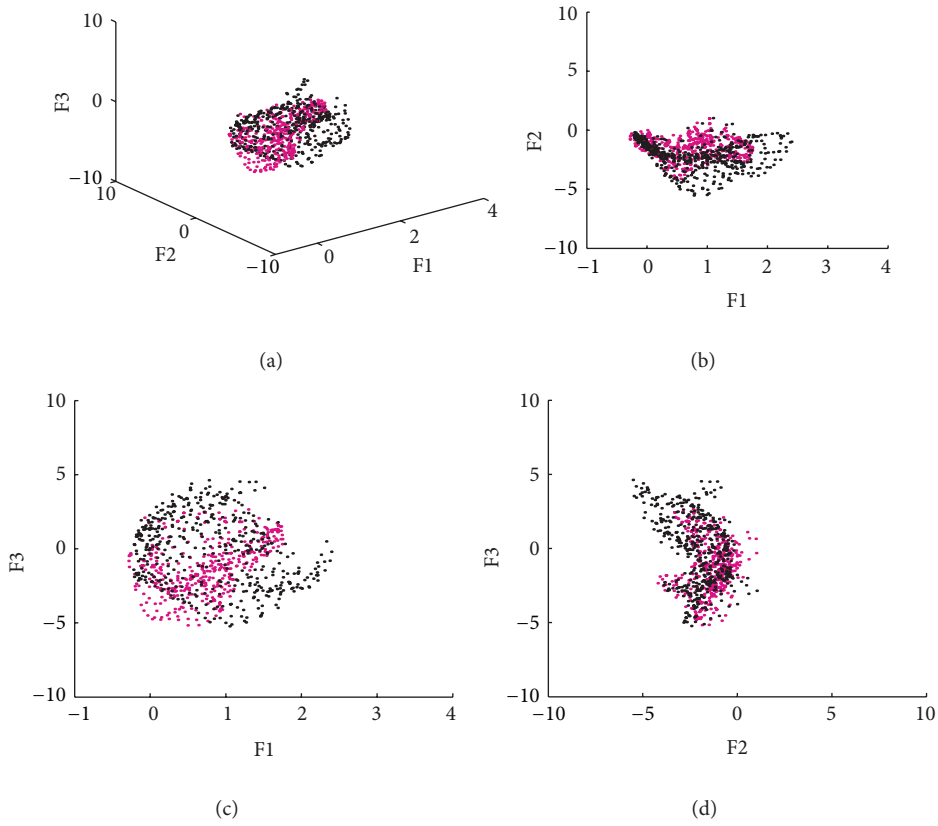


Figure 7: Scatter plot of the 3-MSF (F1, F2, and F3) calculated for a subject with DCM (right ventricle in magenta, left ventricle in black).

cardiac malfunction, still up to 30% of patients do not recover after therapy. Several attempts have been made to improve patient selection and to foresee the successfulness of cardiac resynchronization therapy, depending on the particular dyssynchrony. In a large study, Chung et al. [12] concluded that echocardiographic measures of dyssynchrony are not reliable for this purpose, due to reduced sensitivity and specificity. Furthermore, a complete review by Pavlopoulos and Nihoyannopoulos [10] suggests that an overall approach must be taken to adequately select CRT candidates, where global clinical criteria, in addition to electric and mechanical dyssynchrony measures, must be considered. Van der Wall et al. [11] suggest that, despite the fact that several modalities have been proposed for the noninvasive quantification of LV dyssynchrony, there is no agreement on which technique best predicts response to CRT and that nonechocardiographic imaging techniques, such as ERNA, may provide valuable information for the selection of CRT candidates. Also, Henneman et al. [17] state that phase analysis based on SPECT can lead to an adequate detection of LV dyssynchrony and that nuclear imaging can provide valuable information for the selection of CRT candidates. In summary, the established clinical criteria to consider a subject as a CRT candidate are insufficient to identify those patients that will benefit the most from that treatment [10], justifying the need for an alternative analysis techniques, as the one presented in this work. This method considers information representative of the ventricular contraction dynamics, which is included in the three most significant factors extracted from FADS. It is based on the characterization of a normal contraction pattern, defined from a control population that is used as a reference, against which a "normality index" can be measured. The abnormality in the contraction pattern was globally measured, for two pathologic populations.

An analysis of the 3-MSF scatter plots obtained for the populations studied indicates that the third factor information is necessary to separate left and right ventricular regions, particularly for the abnormal contraction populations, as can be observed in Figures 4, 6, and 7. These findings are in agreement with previously published results [35]. The scatter plots for the DCM patients show an increase in the data number, as well as an overlap between ventricular cavities, as compared to the LBBB population. This can be explained by the fact that DCM is characterized by an important dilation of the ventricles, together with an intrinsically heterogeneous contraction. Therefore, an increase of the information volume and cavities dispersion occurs [38]. Also, as it can be observed in Figures 8, 9, and 10, the adjustment between the normal contraction model and the cardiopathic ventricular behavior is poor, as it was expected.

The comparison of the 3-MSF PDF obtained for the normal population with respect to the behavior of LBBB and DCM groups, using the proposed index, has the advantage of not assuming a specific sequence for the normal

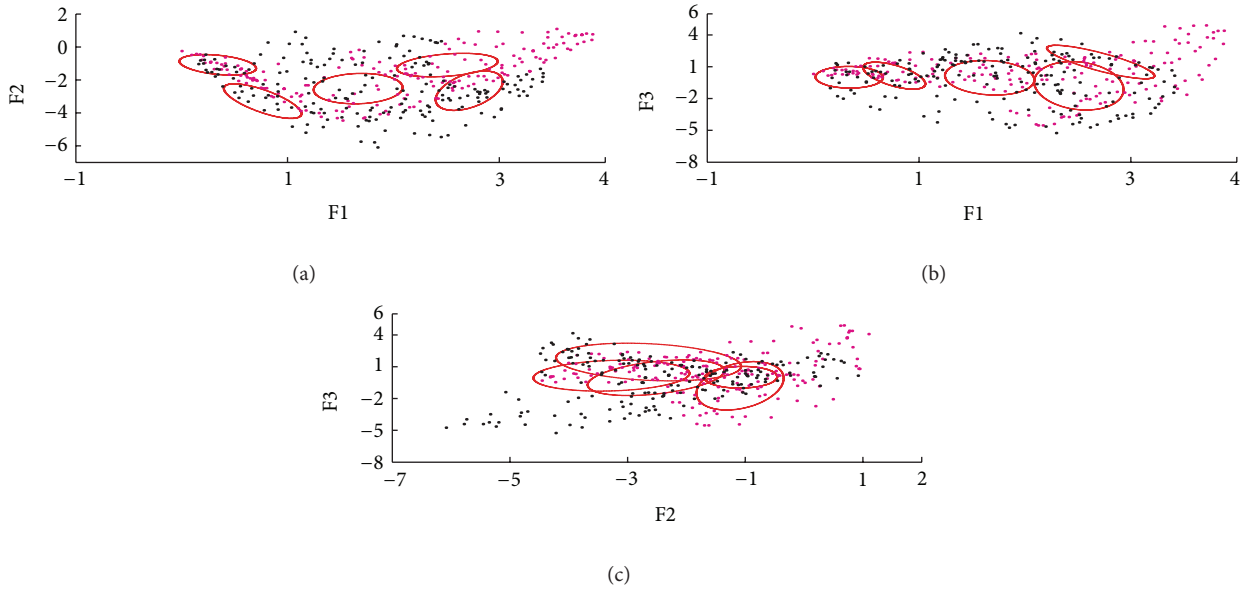


Figure 8: Level curves of the Gaussian functions (in red) superimposed in the dispersion plots of F1, F2, and F3 for a normal subject. Right (magenta) and left (black) ventricle regions are presented.

Table 3: Weight parameters (w), mean values (μ), and covariances (Σ) for the model that best describes the PDF of the 3-MSF of the control population.

Function	Weight (w)	Mean (μ)	Covariance (Σ)
N_1	0.154	$\begin{bmatrix} 0.336 \\ -1.080 \\ 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.131 & -0.077 & -0.003 \\ -0.077 & 0.387 & 0.027 \\ -0.003 & 0.027 & 1.047 \end{bmatrix}$
N_2	0.443	$\begin{bmatrix} 1.661 \\ -2.524 \\ -0.036 \end{bmatrix}$	$\begin{bmatrix} 0.179 & -0.010 & -0.108 \\ -0.011 & 1.352 & 0.788 \\ -0.101 & 0.788 & 2.818 \end{bmatrix}$
N_3	0.139	$\begin{bmatrix} 2.504 \\ -1.096 \\ -0.794 \end{bmatrix}$	$\begin{bmatrix} 0.231 & 0.087 & -0.168 \\ 0.087 & 0.567 & 0.516 \\ -0.168 & 0.516 & 5.219 \end{bmatrix}$
N_4	0.169	$\begin{bmatrix} 0.759 \\ -3.283 \\ 0.160 \end{bmatrix}$	$\begin{bmatrix} 0.165 & -0.425 & -0.026 \\ -0.425 & 1.763 & 0.290 \\ -0.026 & 0.290 & 2.120 \end{bmatrix}$
N_5	0.094	$\begin{bmatrix} 2.710 \\ -2.652 \\ 1.448 \end{bmatrix}$	$\begin{bmatrix} 0.283 & 0.612 & -0.533 \\ 0.612 & 2.479 & -0.641 \\ -0.533 & -0.641 & 3.157 \end{bmatrix}$

contraction pattern and does not depend on the size of the ventricular cavities. It can be observed in Table 4 that the DCM population (inter- and intraventricular asynchrony and EF < 35%) presents a larger and more significant difference with respect to the normal population than the LBBB patients (interventricular asynchrony and normal EF). These findings are in agreement with those reported by Fauchier et al. [26], which conclude that the DCM subjects with inter- and intraventricular dyssynchrony have a greater probability of presenting an adverse cardiac event. Compared to the Fourier phase analysis, the proposed normality index is consistent with the difference obtained with the standard deviation, between normal and pathologic populations. This indicates that the normality index proposed in this work allows the evaluation of the degree of abnormality in ventricular contraction.

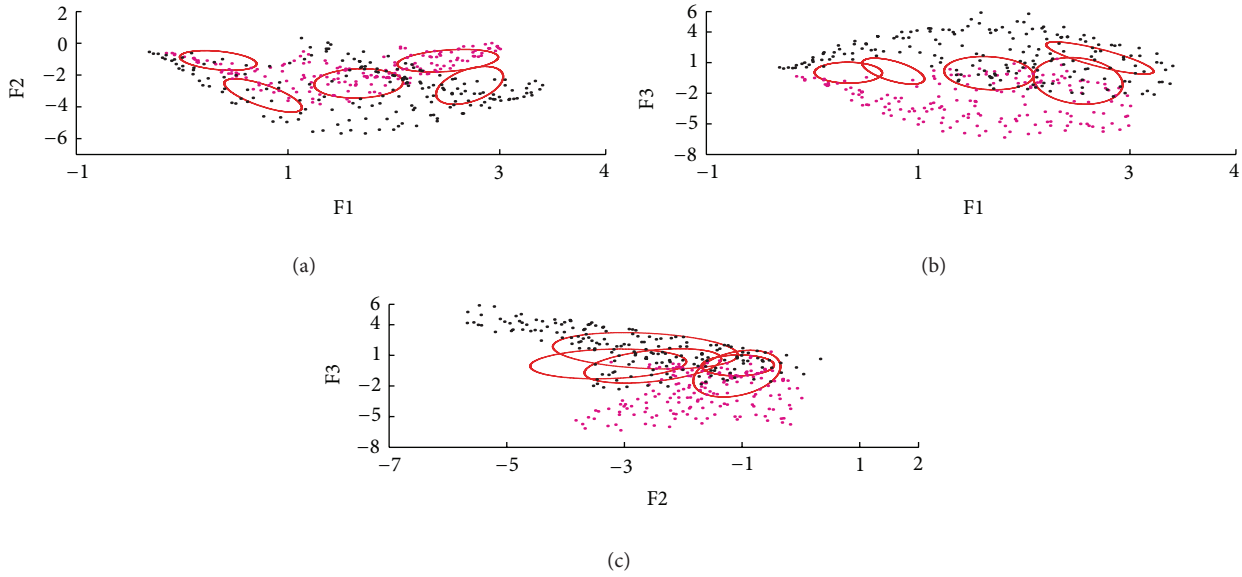


Figure 9: Level curves of the Gaussian functions (in red) superimposed in the dispersion plots of F1, F2, and F3 for a LBBB subject. Right (magenta) and left (black) ventricle regions are presented.

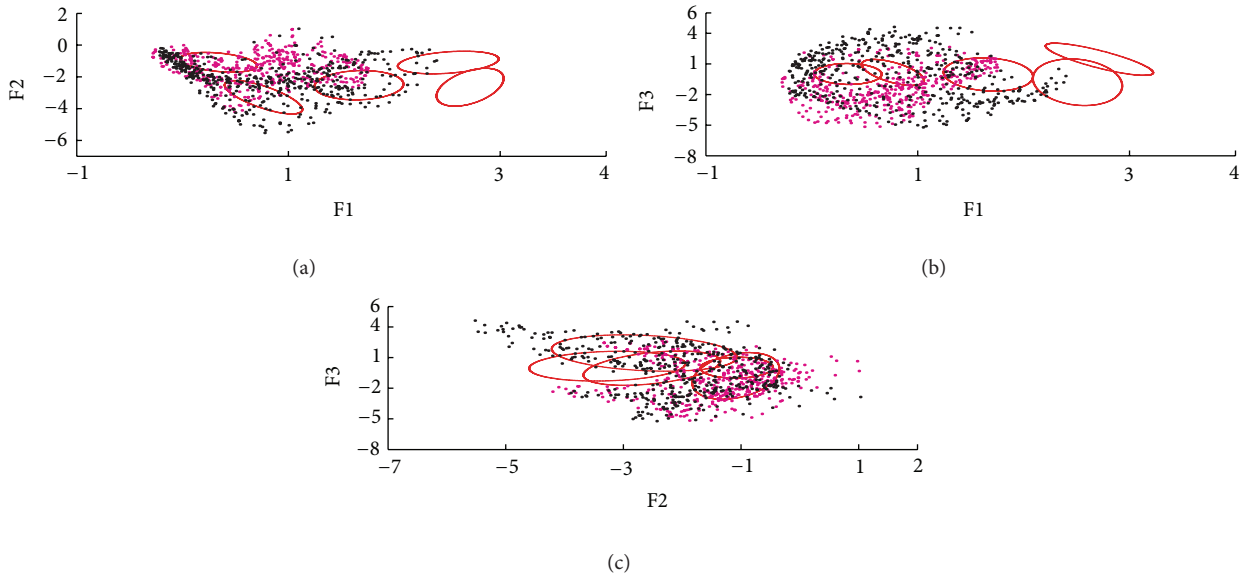


Figure 10: Level curves of the Gaussian functions (in red) superimposed in the dispersion plots of F1, F2, and F3 for a DCM subject. Right (magenta) and left (black) ventricle regions are presented.

5. Conclusions

Due to the reduced sensitivity and specificity of echocardiographic measures to adequately select CRT candidates, ERNA extracted indices seem more reliable in detecting ventricular dyssynchrony. In this work, the probability density function that models a normal ventricular contraction pattern has been defined and validated for a control population. It is based on a thorough analysis of the three most significant dynamic factors obtained from ERNA images in several populations presenting different patterns of ventricular synchrony. Furthermore, a metric to quantify differences between pathologic populations and the reference normal contraction pattern has been proposed; the validation of this likelihood measure was carried out in patients with complete left bundle branch blockage and dilated cardiomyopathy. Both populations presented statistically significant differences in their contraction pattern, compared to the reference model and these differences were more important for the DCM population, due to inter- and intraventricular dyssynchrony, as expected. Compared to the Fourier phase analysis, and particularly with the standard deviation, the

Table 4: Normality indices (see (4)) calculated for Normal, LBBB, and DCM subjects.

	Normal ($n = 8$)	LBBB ($n = 15$)	DCM ($n = 15$)
	FADS		
I_N	1.17 ± 0.12	1.55 ± 0.05*	1.70 ± 0.07*
	Phase analysis		
Mean	127.22	154.13*	156.12
(min, max)	(117.51, 132.63)	(141.79, 170.63)	(128.94, 190.11)
Std. Dev.	12.72	19.36*	46.3405*
(min, max)	(11.19, 14.25)	(18.27, 23.41)	(34.35, 61.24)
Mode	126	150*	156
(min, max)	(118.50, 130.50)	(138, 180)	(132, 198)

* $P < 0.01$ with respect to the control group.

proposed index also detects differences between normal and pathologic groups. Furthermore, this index, together with FADS, was sensitive enough to show contraction pattern differences, which suggests that this analysis can be related to the degree of severity in the ventricular contraction dyssynchrony. Additionally, the use of this index may be promising for the followup of patients under CRT.

In future work, other clinical information should be incorporated, either extracted from EKG or from different imaging modalities, to propose an integrated normality index, to enhance patient selection for CRT. Also, longitudinal studies through different stages of treatment will be necessary to validate the index's capacity to measure asynchrony severity, particularly in patients that have been submitted to cardiac resynchronization therapy.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

This work was supported by CONACyT (Grant no. CB-2006-01-0061657). The authors gratefully acknowledge the assistance of the medical, biomedical, chemical, and technical staff of the Nuclear Cardiology Department at Instituto Nacional de Cardiología Ignacio Chávez.

References

[1] S. Hunt, "ACC/AHA, 2005 guideline update for the diagnosis and management of chronic heart failure in the adult: a report of the American college of cardiology/American heart association task force on practice guidelines (writing committee to update the 2001 guidelines for the evaluation and management of heart failure)," *Journal of the American College of Cardiology*, vol. 46, no. 6, pp. e1–e82, 2005.

[2] D. Lloyd-Jones, R. Adams, M. Carnethon, G. de Simone et al., "Heart disease and stroke statistics 2009 update: a report from the American heart association statistics committee and stroke statistics subcommittee," *Circulation*, vol. 119, pp. e21–e181, 2009.

[3] W. Shamim, D. P. Francis, M. Yousufuddin et al., "Intraventricular conduction delay: a prognostic marker in chronic heart failure," *International Journal of Cardiology*, vol. 70, no. 2, pp. 171–178, 1999.

[4] D. V. Unverferth, R. D. Magorien, M. L. Moeschberger, P. B. Baker, J. K. Fetters, and C. V. Leier, "Factors influencing the one-year mortality of dilated cardiomyopathy," *The American Journal of Cardiology*, vol. 54, no. 1, pp. 147–152, 1984.

[5] D. Farwell, N. R. Patel, A. Hall, S. Ralph, and A. N. Sulke, "How many people with heart failure are appropriate for biventricular resynchronization?" *European Heart Journal*, vol. 21, no. 15, pp. 1246–1250, 2000.

[6] A. Auricchio, C. Stellbrink, S. Sack et al., "Long-term clinical effect of hemodynamically optimized cardiac resynchronization therapy in patients with heart failure and ventricular conduction delay," *Journal of the American College of Cardiology*, vol. 39, no. 12, pp. 2026–2033, 2002.

[7] W. T. Abraham and D. L. Hayes, "Cardiac resynchronization therapy for heart failure," *Circulation*, vol. 108, no. 21, pp. 2596–2603, 2003.

[8] F. A. McAlister, S. Stewart, S. Ferrua, and J. J. J. V. McMurray, "Multidisciplinary strategies for the management of heart failure patients at high risk for admission: a systematic review of randomized trials," *Journal of the American College of Cardiology*, vol. 44, no. 4, pp. 810–819, 2004.

[9] J. G. F. Cleland, J. C. Daubert, E. Erdmann et al., "The effect of cardiac resynchronization on morbidity and mortality in heart failure," *The New England Journal of Medicine*, vol. 352, no. 15, pp. 1539–1549, 2005.

[10] H. Pavlopoulos and P. Nihoyannopoulos, "Recent advances in cardiac resynchronization therapy: echocardiographic modalities, patient selection, optimization, non-responders-all you need to know for more efficient CRT," *International Journal of Cardiovascular Imaging*, vol. 26, no. 2, pp. 177–191, 2010.

[11] E. E. van der Wall, M. J. Schalij, and J. J. Bax, "Cardiac resynchronization therapy; evaluation by advanced imaging techniques," *International Journal of Cardiovascular Imaging*, vol. 26, no. 2, pp. 199–202, 2010.

[12] E. S. Chung, A. R. Leon, L. Tavazzi et al., "Results of the predictors of response to crt (prospect) trial," *Circulation*, vol. 117, no. 20, pp. 2608–2616, 2008.

[13] K. C. Bilchick, V. Dimaano, K. C. Wu et al., "Cardiac magnetic resonance assessment of dyssynchrony and myocardial scar predicts function class improvement following cardiac resynchronization therapy," *JACC: Cardiovascular Imaging*, vol. 1, no. 5, pp. 561–568, 2008.

[14] I. K. Rüssel, J. Dijk, S. A. Kleijn et al., "Relation between three-dimensional echocardiography derived left ventricular volume and MRI derived circumferential strain in patients eligible for cardiac resynchronization therapy," *International Journal of Cardiovascular Imaging*, vol. 25, no. 1, pp. 1–11, 2009.

[15] J. J. M. Westenberg, H. J. Lamb, R. J. van der Geest et al., "Assessment of left ventricular dyssynchrony in patients with conduction delay and idiopathic dilated cardiomyopathy. Head-to-head comparison between tissue doppler imaging and velocity-encoded magnetic resonance imaging," *Journal of the American College of Cardiology*, vol. 47, no. 10, pp. 2042–2048, 2006.

[16] A. Roguin, J. Schwitter, C. Vahlhaus et al., "Magnetic resonance imaging in individuals with cardiovascular implantable electronic devices," *Europace*, vol. 10, no. 3, pp. 336–346, 2008.

[17] M. M. Henneman, E. E. van der Wall, C. Ypenburg et al., "Nuclear imaging in cardiac resynchronization therapy," *Journal of Nuclear Medicine*, vol. 48, no. 12, pp. 2001–2010, 2007.

[18] T. F. J. Wackers, H. J. Berger, and D. E. Johnstone, "Multiple gated cardiac blood pool imaging for left ventricular ejection fraction: validation of the technique and assessment of variability," *The American Journal of Cardiology*, vol. 43, no. 6, pp. 1159–1166, 1979.

[19] K. A. Williams, "Measurement of ventricular function with scintigraphic techniques—part II: ventricular function with gated techniques for blood pool and perfusión imaging," *Journal of Nuclear Cardiology*, vol. 12, no. 2, pp. 208–215, 2005.

[20] E. Botvinick, R. Dunn, M. Frais et al., "The phase image: its relationship to patterns of contraction and conduction," *Circulation*, vol. 65, no. 3, pp. 551–560, 1982.

[21] J. W. O'Connell, C. Schreck, M. Moles et al., "A unique method by which to quantitate synchrony with equilibrium radionuclide angiography," *Journal of Nuclear Cardiology*, vol. 12, no. 4, pp. 441–450, 2005.

[22] K. J. Nichols, A. van Tosh, P. de Bondt, S. R. Bergmann, C. J. Palestro, and N. Reichek, "Normal limits of gated blood pool SPECT count-based regional cardiac function parameters," *International Journal of Cardiovascular Imaging*, vol. 24, no. 7, pp. 717–725, 2008.

[23] J. Chen, E. V. Garcia, R. D. Folks et al., "Onset of left ventricular mechanical contraction as determined by phase analysis of ECG-gated myocardial perfusion SPECT imaging: development of a diagnostic tool for assessment of cardiac mechanical dyssynchrony," *Journal of Nuclear Cardiology*, vol. 12, no. 6, pp. 687–695, 2005.

[24] J. F. Toussaint, A. Peix, T. Lavergne et al., "Reproducibility of the ventricular synchronization parameters assessed by multiharmonic phase analysis of radionuclide angiography in the normal heart," *International Journal of Cardiovascular Imaging*, vol. 18, no. 3, pp. 187–194, 2002.

[25] W. F. Kerwin, E. H. Botvinick, J. W. O'Connell et al., "Ventricular contraction abnormalities in dilated cardiomyopathy: effect of biventricular pacing to correct interventricular dyssynchrony," *Journal of the American College of Cardiology*, vol. 35, no. 5, pp. 1221–1227, 2000.

[26] L. Fauchier, O. Marie, D. Casset-Senon, D. Babuty, P. Cosnay, and J. P. Fauchier, "Interventricular and intraventricular dyssynchrony in idiopathic dilated cardiomyopathy: a prognostic study with Fourier phase analysis of radionuclide angioscintigraphy," *Journal of the American College of Cardiology*, vol. 40, no. 11, pp. 2022–2030, 2002.

[27] L. Fauchier, O. Marie, D. Casset-Senon, D. Babuty, P. Cosnay, and J. P. Fauchier, "Reliability of QRS duration and morphology on surface electrocardiogram to identify ventricular dyssynchrony in patients with idiopathic dilated cardiomyopathy," *The American Journal of Cardiology*, vol. 92, no. 3, pp. 341–344, 2003.

[28] J. F. Toussaint, T. Lavergne, K. Kerrou et al., "Basal asynchrony and resynchronization with biventricular pacing predict long-term improvement of LV function in heart failure patients," *Pacing and Clinical Electrophysiology*, vol. 26, no. 9, pp. 1815–1823, 2003.

[29] L. Fauchier, V. Eder, D. Casset-Senon et al., "Segmental wall motion abnormalities in idiopathic dilated cardiomyopathy and their effect on prognosis," *The American Journal of Cardiology*, vol. 93, no. 12, pp. 1504–1509, 2004.

[30] R. Dauphin, E. Nonin, L. Bontemps et al., "Quantification of ventricular resynchronization reserve by radionuclide phase analysis in heart failure patients: a prospective long-term study," *Circulation: Cardiovascular Imaging*, vol. 4, no. 2, pp. 114–121, 2011.

[31] F. Frouin, A. Delouche, H. Raffoul, H. Diebold, E. Abergel, and B. Diebold, "Factor analysis of the left ventricle by echocardiography (FALVE): a new tool for detecting regional wall motion abnormalities," *European Journal of Echocardiography*, vol. 5, no. 5, pp. 335–346, 2004.

[32] F. Cavailloles, J. P. Bazin, D. Pavel et al., "Comparison between factor analysis of dynamic structures and Fourier analysis in detection of segmental wall motion abnormalities: a clinical evaluation," *International Journal of Cardiac Imaging*, vol. 11, no. 4, pp. 263–272, 1995.

[33] D. C. Barber, "The use of principal components in the quantitative analysis of gamma dynamic studies," *Physics in Medicine and Biology*, vol. 25, no. 2, pp. 283–292, 1980.

[34] R. Di Paola, J. P. Bazin, F. Aubury et al., "Handling of dynamic sequences in nuclear medicine," *IEEE Transactions on Nuclear Science*, vol. 29, no. 4, pp. 1310–1321, 1982.

[35] L. Jiménez-Ángeles, R. Valdés-Cristerna, E. Vallejo, D. Bialostozky, and V. Medina-Bañuelos, "Factorial phase analysis of ventricular contraction using equilibrium radionuclide angiography images," *Biomedical Signal Processing and Control*, vol. 4, no. 2, pp. 149–161, 2009.

[36] C. M. Bishop, *Continuous Latent Variables in Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.

[37] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximun likelihood from imcomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.

[38] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[39] C. Fraley and A. E. Raftery, "Model-based methods of classification: using the mclust software in chemometrics," *Journal of Statistical Software*, vol. 18, no. 6, pp. 1–13, 2007.

[40] R. Ihaka and R. Gentleman, "R: a language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299–314, 1996.

[41] D. P. Miller, N. McCabe, C. Pye et al., "Natural history of isolated bundle branch block," *The American Journal of Cardiology*, vol. 77, no. 14, pp. 1185–1190, 1996.

[42] Criteria Committee, New York Heart Association, *Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels*, Little, Brown and Company, Boston, Mass, USA, 9th edition, 1994.

[43] D. G. William and V. Fuster, "Idiopathic dilated cardiomyopathy," *The New England Journal of Medicine*, vol. 331, pp. 1564–1575, 1994.

[44] R. Schoeller, D. Andresen, P. Buttner, K. Oezcelik, G. Vey, and R. Schroder, "First- or second-degree atrioventricular block as a risk factor in idiopathic dilated cardiomyopathy," *The American Journal of Cardiology*, vol. 71, no. 8, pp. 720–726, 1993.

[45] K. J. Nichols, S. Bacharach, S. Bergmann et al., "ASNC imaging guidelines for nuclear cardiology procedures: instrumentation quality assurance and performance," *Journal of Nuclear Cardiology*, vol. 14, no. 6, pp. e61–e78, 2007.

[46] R. J. Bunder, I. Haluszcynski, and H. Langhammer, "In vivo/in vitro labeling of red blood cells with Tc-99m," *European Journal of Nuclear Medicine and Molecular*, vol. 8, pp. 218–225, 1983.

[47] *UltraTag RBC*, Mallinckrodt Medical, St Louis, Mo, USA, 1995.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 213794, 10 pages
http://dx.doi.org/10.1155/2013/213794

Research Article

Study of the Effect of Breast Tissue Density on Detection of Masses in Mammograms

A. García-Manso, C. J. García-Orellana, H. M. González-Velasco, R. Gallardo-Caballero, and M. Macías-Macías

Pattern Classification and Image Analysis Group, University of Extremadura, Avenida de Elvas s/n, Badajoz, 06006 Extremadura, Spain

Correspondence should be addressed to A. García-Manso; antonio@capi.unex.es

Received 4 October 2012; Accepted 21 February 2013

Academic Editor: Angel García-Crespo

One of the parameters that are usually stored for mammograms is the BI-RADS density, which gives an idea of the breast tissue composition. In this work, we study the effect of BI-RADS density in our ongoing project for developing an image-based CAD system to detect masses in mammograms. This system consists of two stages. First, a blind feature extraction is performed for regions of interest (ROIs), using Independent Component Analysis (ICA). Next, in the second stage, those features form the input vectors to a classifier, neural network, or SVM classifier. To train and test our system, the Digital Database for Screening Mammography (DDSM) was used. The results obtained show that the maximum variation in the performance of our system considering only prototypes obtained from mammograms with a concrete value of density (both for training and test) is about 7%, yielding the best values for density equal to 1, and the worst for density equal to 4, for both classifiers. Finally, with the overall results (i.e., using prototypes from mammograms with all the possible values of densities), we obtained a difference in performance that is only 2% lower than the maximum, also for both classifiers.

1. Introduction

Several factors can affect the composition of breast tissue. The increase or decrease of the breast gland is part of the normal physiological changes that occur in the breast and usually occurs in both breasts simultaneously. These changes may be caused by hormonal fluctuations (natural or synthetic) including menarche, pregnancy, breastfeeding, or menopause. The increase in glandularity also depends on the woman's genetic predisposition. In young women, normally, the breast is composed mostly of glandular tissue and very little fat. And although this composition varies depending on age, it is possible to find older women with extremely dense breasts, that is, consisting mostly of glandular tissue and not fat. Weight gain or loss also increases or decreases the fat content of the breast and therefore also affects the breast glandularity [1].

The composition of breast tissue is defined by the BI-RADS parameter called "density" [2], which can have four possible values (1–4) explained in Table 1.

The degree of difficulty of analyzing a mammogram depends on the nature of the breast tissue, as can be seen in Figure 1. In these two mammograms, the different nature of the tissue predominant in each one is clearly distinguishable. As can be seen, it is very easy to locate the lesion in the figure on the left, which corresponds to a 71-years-old woman and has a density equal to 1, whereas it is much more difficult to analyze and locate the lesion in the mammogram on the right, corresponding to a 41-years-old woman with a density equal to 4. This example suggests that the density may be a factor limiting the sensitivity which can be reached when analyzing a mammography (both for radiologists or CAD systems). Several analyses can be found showing that the majority of cancer cases discarded in screening mammographies correspond to dense mammary gland (density equal to 3 or 4) [3–5]. We can also find in the literature works as [6], in which the impact of BI-RADS density on CAD systems is studied, in particular on the SecondLook CAD system (version 4.0) developed by the company iCAD. Finally, there are other

Table 1: Meaning of the BI-RADS density.

BI-RADS density	
Density value	Description
1	Breast tissue mainly fatty
2	Scattered fibroglandular densities
3	Breast tissue heterogeneously dense
4	Breast tissue extremely dense

studies such as [7] that incorporate the information provided by this parameter for the development of their algorithms to detect masses in mammograms. In this work, we studied how BI-RADS density affects our mass detection system, which consists of two stages. In the first one, a blind feature extraction is performed over ROIs, using ICA as the main technique. Next, in the second stage, those features are used as inputs to a neural classifier which determines whether the ROI includes a mass. The system is described in detail in the next two sections.

The rest of our paper is organized as follows. Section 2 describes the general methods used for the generation of prototypes, feature extraction, and classification. Next, Section 3 includes a description of the system structure and operation, and also of the experiments devised. In Section 4, the most significant obtained results are described, while Section 5 presents the main conclusions of the work.

2. Methods

In this section, we present the techniques used in this study for the generation and selection of prototypes, for feature extraction tasks, and for classification. We are going to review these methods in the following subsections.

2.1. Data and Prototype Creation. In the literature, one can find various proposals focused on the detection and segmentation of masses on mammograms, such as those reviewed in [8], but it is usually difficult to compare the results of different studies addressing both the detection and diagnosis of masses. The main problem is the use of proprietary databases of small size, or, if using a public database, the use of selected, unspecified cases. Horsch [9] analyzes recent studies in mammography CAD and concludes that, in view of the observed variability in the datasets used, currently the only mammography database that is public and sufficiently large to allow a meaningful and reproducible evaluation of a CAD system is the Digital Database for Screening Mammography (DDSM) [10].

The DDSM is a resource available to the mammographic image analysis research community and contains a total of 2,620 cases. Each case provides four screening views: mediolateral oblique (MLO) and craniocaudal (CC) projections of left and right breasts. Therefore, the database has a total of 10,480 images. Cases are categorized in four major groups: *normal*, *cancer*, *benign,* and *benign without callback*. All cases in the DDSM were reported by experienced radiologists providing various BI-RADS parameters (density, assessment, and subtlety), BI-RADS abnormality description, and proven pathology. For each abnormality identified (within which masses are included), the radiologists draw free form digital curves defining ground truth regions. We consider these regions to define squared "regions of interest" (ROIs) that will be used as prototypes of mass. Apart from the previous data, each DDSM case includes additional information such as patient age, date of study, and digitization or digitizer's brand, though we have not used it in this work.

The DDSM database contains 2,582 images that contain an abnormality identified as mass, whether benign or malignant. Some of them were located on the border of the mammograms and could not be used (see the following paragraph, dedicated to ROIs). Consequently, only 2,324 prototypes could be considered, namely, those which might be taken centered in a square without stretching. Some mass prototype examples are shown in Figure 2.

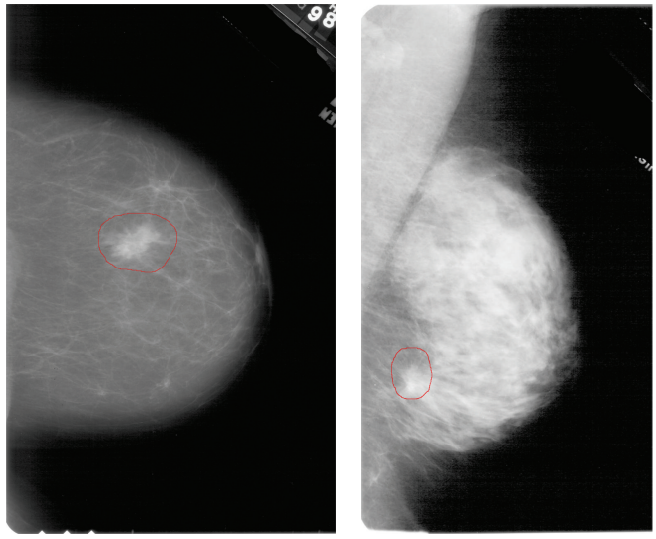
Regions of Interest. Ground truth regions for abnormalities are defined in the database by a chain code which generates a free hand closed curve. We use the chain code to determine the smallest square region of the mammogram that includes the manually defined region. Therefore, if the mass is located near one edge of the mammogram, this procedure may not be able to obtain a squared region from the image, and the mass is discarded as a valid prototype. Figure 3 shows an example of the ground truth region coded by the radiologist (solid line) and the area to be used as ROI (purple box). On the other hand, the prototypes of normal tissue were selected randomly from the normal mammograms. This normal tissue prototypes were caught originally with sizes randomly ranging from the smallest to the largest of the sizes found in the DDSM for masses.

The generated regions have different sizes but the selected image feature extractor needs to operate on regions with the same size, so we need to reduce the size of the selected regions to common sizes. The reduction of ROIs to a common size has demonstrated to preserve mass malignancy information [11–13]. To determine the optimum region size, we considered two sizes for the experiments: 32×32, 64×64 pixels. The process of resizing was carried out using the bilinear interpolation algorithm provided by the OpenCV library [14].

2.2. Feature Extraction. As we commented above, we used Independent Component Analysis (ICA) [15] as blind feature extraction method. The objective of the method is to obtain an appropriate functions basis, derived from prototype ROIs (including masses and normal tissue), so that we can represent the texture and characteristics of each ROI from the breast images as an expansion in this basis (Figure 4), where the coefficients of this expansion ($\mathbf{s}_i$) are the input vectors to the classifiers (i.e., the "features" describing the ROIs).

The added value of our approach, compared to other methods that use some generic functions, is that our basis should be more specific for our problem, since it is obtained using a selection of the images to be classified.

Independent Component Analysis. Independent Component Analysis (ICA) defines a generative model of the observed



(a) (b)

Figure 1: The left image shows the RCC view (right craniocaudal) of the case 1468 in USF's DDSM database that corresponds to a woman of 71 years, to which the radiologist assigned a density equal to 1. The right image shows the RMLO view (right mediolateral oblique) of the case 1985 in the same database corresponding to a woman of 41 years, and density equal to 4.

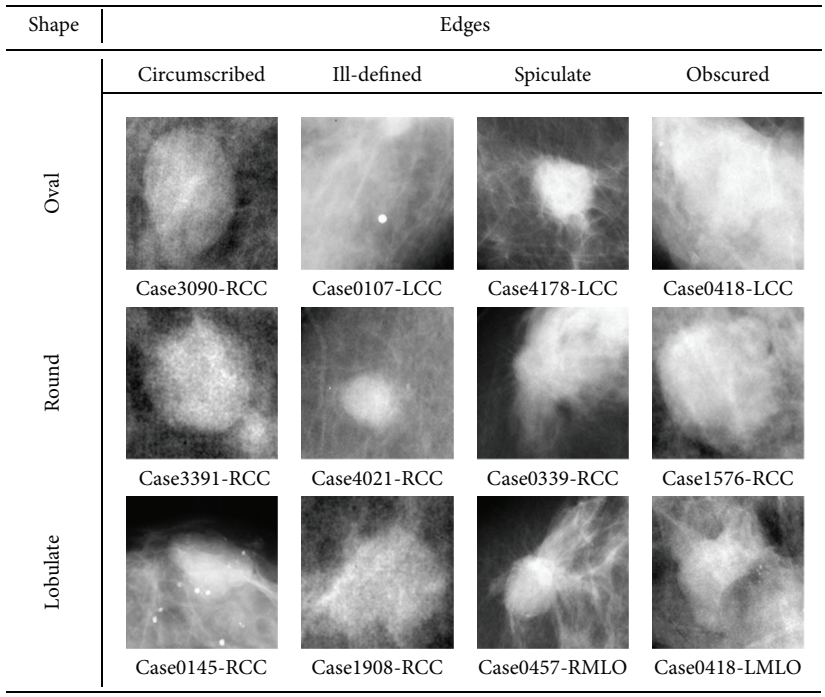


Figure 2: Examples of masses for each combination of shape and margins. Each ROI image has been resized to a common size of 128 × 128 pixels. Case name and view are located at the bottom of each ROI.

multivariate data, typically given as a sample database. In this model, it is assumed that the data are linear combinations of some unknown latent variables, and the system by which are combined is also unknown. It is assumed that the latent variables are non-Gaussian and mutually independent, and they are called independent components of the observed data. These independent components, also called sources or factors, can be determined by ICA. ICA is related to Principal Component Analysis (PCA) [16] since, before applying the ICA method itself, it is advisable to make a dimension reduction or feature extraction of the original input vectors which can be done using PCA. The data analyzed by ICA can come from many different types of fields including digital images. In many cases, the data comes from a set of parallel

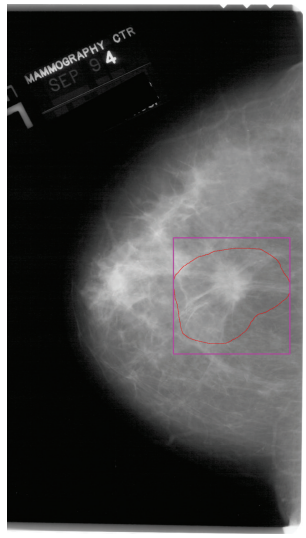


Figure 3: Ground truth region was defined by radiologist (red solid line) and was considered ROI (purple box) on a DDSM mammogram.

signals or time series, being used in this case the term "Blind Source Separation" (BSS) to define these problems.

In that sense, if we suppose that we have n signals, the objective is to expand the signals registered by the sensors ($\mathbf{x}_i$) as a linear combination of n sources ($\mathbf{s}_j$), in principle unknown as follows:

$$\mathbf{x}_i = \sum_{j=1}^{n} a_{ij}\mathbf{s}_j. \tag{1}$$

The goal of ICA is to estimate the mixing matrix $\mathbf{A} = (a_{ij})$, in addition to the sources $\mathbf{s}_j$. One can use this technique for feature extraction since the components of $\mathbf{X}$ can be regarded as the characteristics representing the objects (patterns) [15].

2.3. Classification Algorithm. In our system, the classification algorithm has the task of learning from data. An excessively complex model will usually lead to poorly generalizable results. It is advisable to use at least two independent sets of patterns in the learning process: one for training and another for testing. In the present work, we use three independent sets of patterns: one for training, one to avoid overtraining (validation set), and another for testing [17]. For the classification, we have used Multilayer Perceptron (MLP) [18] and SVM classifiers [19]. We have chosen these two techniques because they are widely used in classification and detection of breast cancer, as can be seen in the works listed in several reviews as [9] and in [20]. Also, to do a more rigorous study as is shown in [21], we could have tested with other techniques and other quality metrics that are also widely used in classification and regression problems, although they may not be as common in works found on detection and classification of breast cancer.

2.3.1. Neural Networks. We implement MLP with a single hidden layer, and a variant of the Back-Propagation algorithm termed Resilient Back-Propagation (Rprop) [22] to adjust the weights. This last is a local adaptive learning scheme performing supervised batch-learning in a multilayer perceptron which converges faster than the standard BP algorithm. The basic principle of Rprop is to eliminate the negative effect of the size of the partial derivative on the update process. As a consequence, only the sign of the derivative is considered in indicating the direction of the weight update [22]. The function library of the Stuttgart Neural Network Simulator environment [23] was used to generate and train the NN classifiers. To avoid local minimum during the training process, each setting was repeated four times, changing the initial weights in the net at random. Furthermore, the number of neurons in the hidden layer was allowed to vary between 50 and 650 in steps of 50.

2.3.2. Support Vector Machines. As with MLP, the goal of using an SVM is to find a model (based on the training prototypes) which is able to predict the class membership of the test subset's prototypes based on the value of their characteristics. Given a labeled training set of the form $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \ldots, l$ where $\mathbf{x}_i \in \Re^n$ and $\mathbf{y} \in \{1, -1\}^l$, the SVM algorithm involves solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}\in\Re^d, b, \xi_i \in \Re^+} \quad & \|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i\left(\mathbf{w}^T\phi\left(\mathbf{x}_i\right) + b\right) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{2}$$

In this algorithm, the training vectors $\mathbf{x}_i$ are projected onto a higher-dimensional space than the original. The final dimension of this space depends on the complexity of the input space. Then the SVM finds a linear separation in terms of a hyperplane with a maximal (and hence optimal) margin of separation between classes in this higher dimensional space.

In the model, C ($C > 0$) is a regularization or penalty parameter to control the error, d is the final dimension of the projection space, $\mathbf{w}$ is the normal to the hyperplane (also known as the weights vector), and b is the bias. The parameter ξ is introduced to allow the algorithm a degree of flexibility in fitting the data, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ is a kernel function to project the input data onto to a higher dimensional space. We used the LibSVM [24] library with a radial basis function (RBF: $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2), \gamma > 0$) as kernel function. To find the optimal configuration of the parameters in the algorithm, γ was allowed to vary like $2^{-5} < \gamma < 2^3$ in steps of 0.5 for the exponent, and the penalty parameter C between 2^{-5} and 2^{10} also in steps of 0.5 for the exponent.

3. Outline of the Process

In this section, we provide an overview of the structure of our system, describing the main steps required to configure the

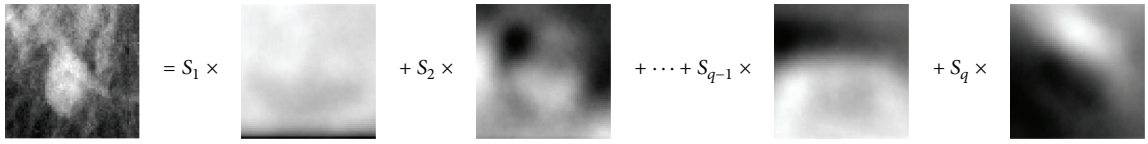


Figure 4: Decomposition of the image using an ICA basis.

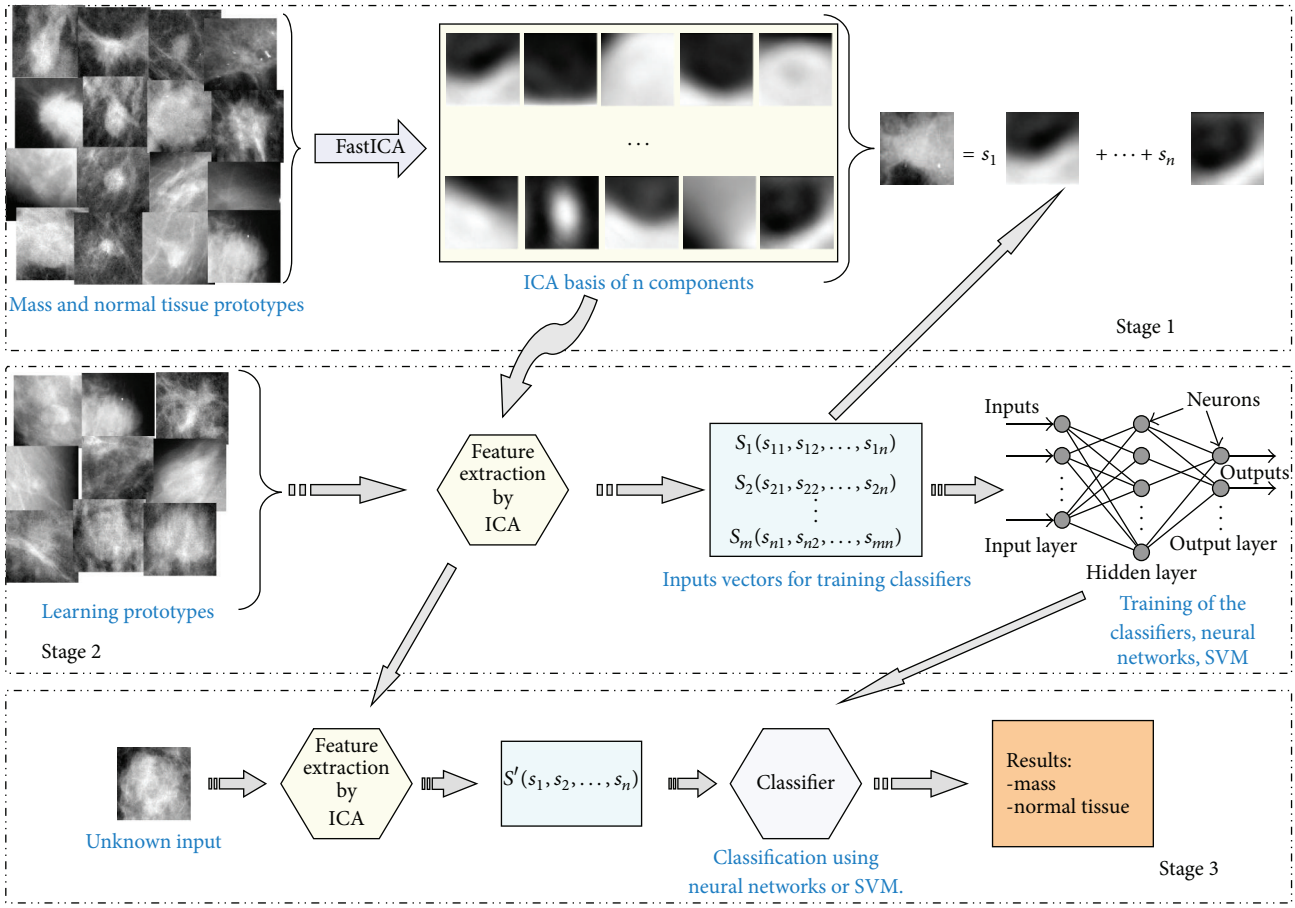


Figure 5: Overview of the system proposed.

system to discriminate prototypes of masses from prototypes of normal breast tissue.

3.1. System Description. We provide an overview of our system's structure, describing the main steps required to configure the system in order to discriminate ROIs corresponding to masses from ROIs corresponding to normal tissue. In addition, we will present the experiments devised to determine how the performance of these classifiers is affected by the breast density, that is associated with each mammography (and, therefore, with each ROI).

The main scheme that summarizes in a more graphical form all phases of this work is represented in Figure 5. In the first stage, the prototypes of masses are obtained as was explained in Section 2.1. Then the FastICA algorithm [25, 26] is applied to obtain the ICA basis (the ICA-based feature extractor), with the *log cosh* function being used to approximate the neg-entropy. These bases are generated with different configurations, different numbers of components, and using prototypes of different sizes. The second stage uses this generated basis to obtain the training sets and to train and test the classifiers. Finally, in the third stage, the test subset, which contains input vectors not used in the optimization of the classifiers, is used to provide performance results of our system.

3.2. System Optimization. To determine the optimal configuration of the system, various ICA bases were generated to extract different numbers of features (from 10 to 65 in steps of 5) from the original patches, and operating on patches of the different sizes noted above (32 × 32 and 64 × 64 pixels).

The training process consisted of two stages—first training the NN classifiers, and then the SVM classifiers. The results thus obtained on the test subsets in a 10-fold cross

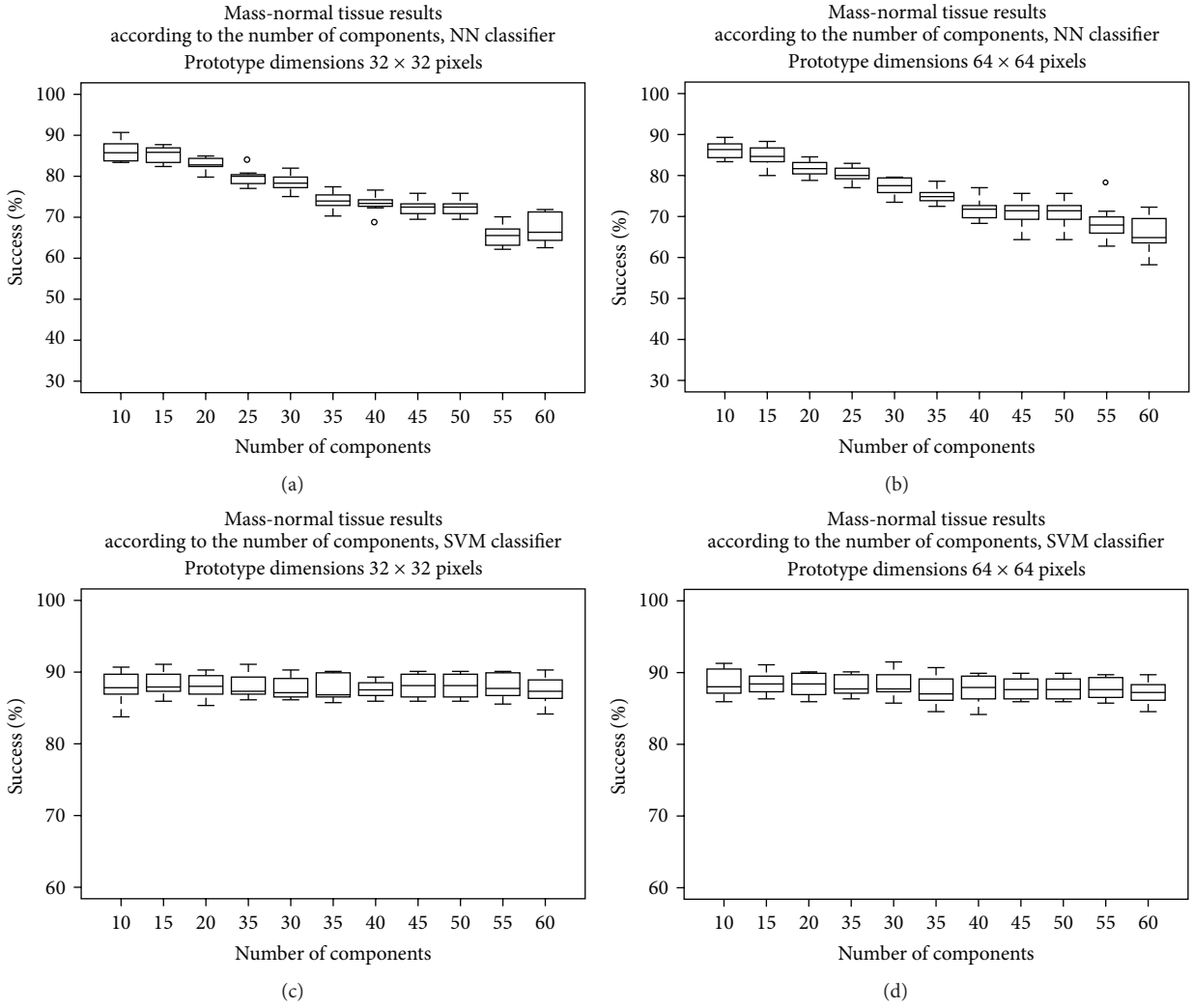


Figure 6: Choosing the best configuration for the feature extractor. The top row shows the results when using an NN classifier, and the bottom row shows the results for an SVM classifier. In both cases, prototypes of 32 × 32 are in the first column, and of 64 × 64 in the second column.

validation scheme are shown in Figure 6. This allowed us to find the optimal configuration of the feature extractor.

The study was done with a total of 5052 prototypes: 1197 of malignant masses, 1133 of benign masses, and 2722 of normal tissue.

We found that the optimal ICA-based feature extractor configuration for an NN classifier was a feature extractor that operated on prototypes of 64 × 64 pixels, extracting 10 components (average success rate 86.33%), and for an SVM classifier was a feature extractor that also operated on prototypes of 64 × 64 pixels, extracting 15 components (average success rate 88.41%). The results to be presented in the following section were obtained using these optimal configurations.

3.3. Experiments. To determine how the density associated to each mammography (and, therefore, to each ROI) could affect the performance of our system, we carried out five experiments. In each of the experiments we made the same tests, but with different sets of prototypes: first with all the available prototypes (one experiment), and then with prototypes obtained from mammograms with a given value of density (four experiments).

For each of the experiments, a 30-fold cross validation scheme was used. In this process, 30 partitions of the data set are generated randomly, and, iteratively, one partition is reserved for test, and the remaining 29 are used for training and validation (80% of the prototypes for training and 20% for validation). As a result we have 30 performance values that can be studied statistically.

Finally, to analyze the performance and compare results, ROC curves [27] have been generated for each experiment. To this end, the threshold applied to the output neuron of the classifier (in order to decide if the prototype being classified is mass or normal tissue) is swept, and the ratios of true and false positives are calculated. As a performance parameter, the "area under curve" (AUC) was used.

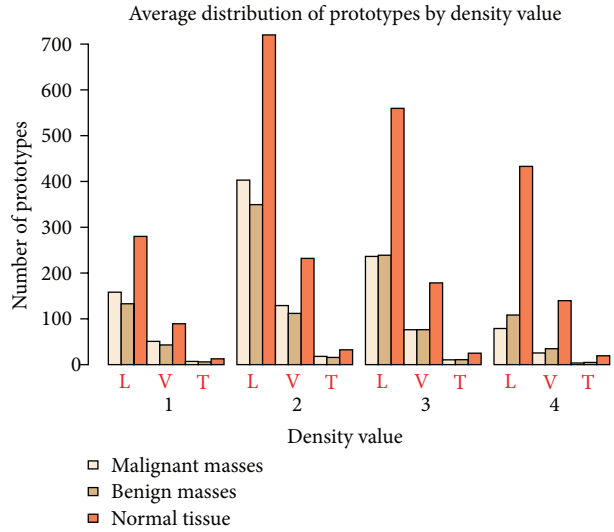


Figure 7: Average number of prototypes of malignant and benign masses and normal tissue divided into training (L), validation (V), and test (T) sets, distributed by density value.

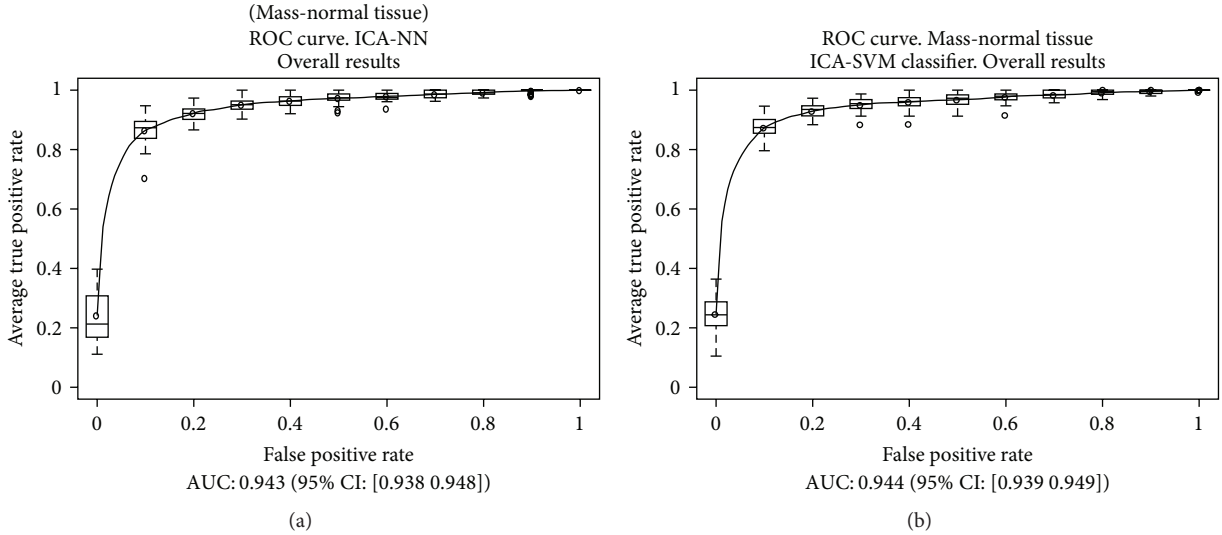


Figure 8: Results obtained over the test subsets, considering all the prototypes.

Regarding the prototypes, Table 2 shows the average number of "normal breast tissue," "benign mass," and "malignant mass" prototypes for each of the subsets (training, validation, and test), and calculated over the 30 "trainings of the classifier" that are made in the 30-fold cross validation scheme. These average values are shown for the overall experiment, and for the experiments with a given value of density. In the process of selection of the prototypes, no account was taken of the pathology of them. But, as can be seen, this selection process yields always a balanced distribution of the mean number of prototypes in each subset. On average, about 73% of malignant prototypes were included in the training sets, 23% in the validation sets, and 3% in the test sets. For the case of the benign prototypes, around 73% were included in the training sets, a 23% in the validation sets, and 3% in the test sets. And finally, in the case of normal prototypes, about 73% were included in the training sets, 23% in the sets of validation, and a 3% in the test sets. Therefore, if we only consider the overall data, there seems to be no clear trend which suggests that the prototypes selected in any of the ranges of density have a greater or lesser likelihood of being mass or normal tissue. However, when we analyze particular density values, differences are observed in the number of prototypes for each class that may be significant.

In Figure 7, it can be seen that the prototypes of malignant and benign masses prototypes are quite different from the number of prototypes of normal tissue in some cases. For a density value equal to 3, this sum is always significantly lower than the number of normal tissue prototypes. For example, in the training subset this sum is equal to 475.2 and the number of normal tissue prototypes is equal to 559.6. Therefore, there is a difference of 15%. Moreover, this difference is much

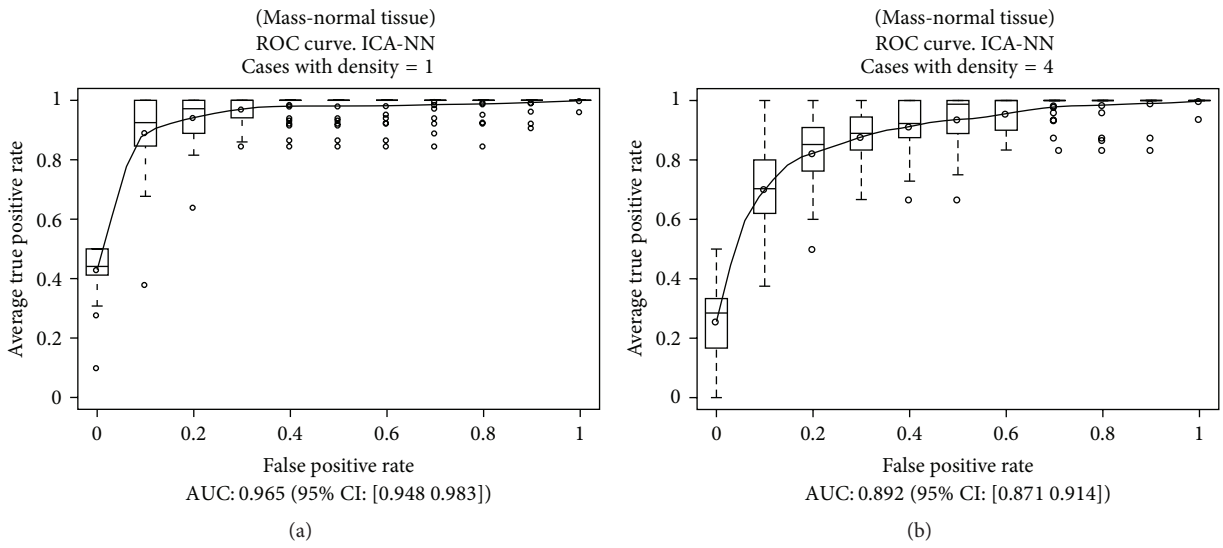


Figure 9: Results obtained over the test subsets, considering NN classifiers and the cases of density 1 and 4.

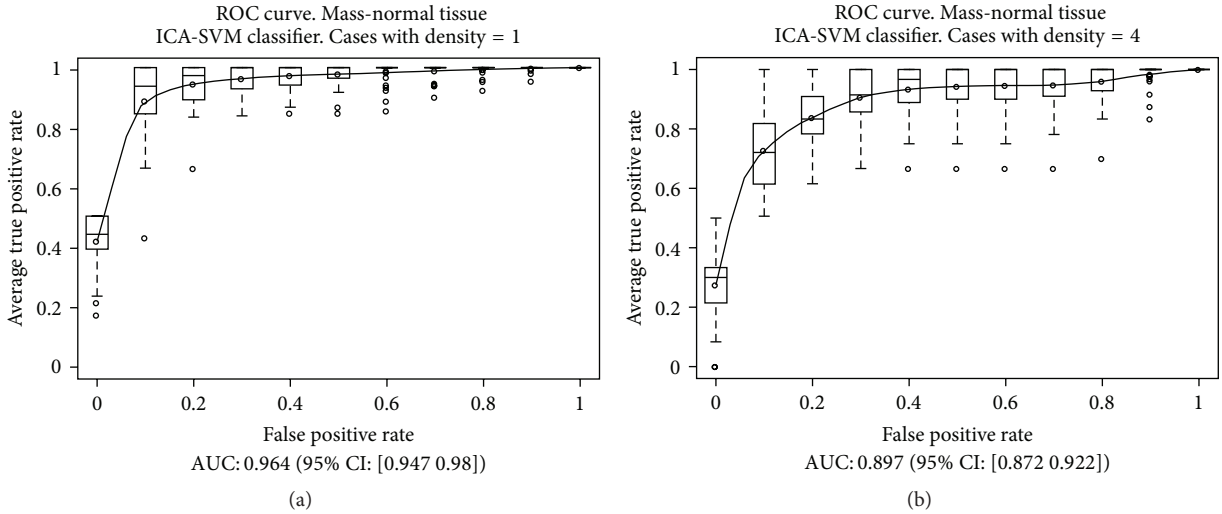


Figure 10: Results obtained over the test subsets, considering SVM classifiers and the cases of density 1 and 4.

more significant for a density value equal to 4, where, for the training subset, the sum of malignant and benign masses is equal to 187.2 and the number of normal tissue prototypes is 432.9, being, therefore, the difference equal to 57%. In contrast, for density values equal to 1 and 2 these differences are just only a 3% and 4%, respectively, favorable to the number of mass prototypes.

4. Results

As we stated above, our main interest in this paper is to evaluate the dependence presented by our system with the composition of breast tissue, determined by the BI-RADS density parameter. For this study, we have considered all those prototypes of masses in the DDSM for which a square shape could be obtained by determining the smallest squared region that includes the complete area marked by the radiologist, and always without resizing. As we commented before, the distribution of prototypes is shown in Table 2 and in Figure 7. We must point out that the relative number of prototypes of each class is very different depending on the density value. Particularly, for a density value of 4, the difference between mass (malignant and benign) prototypes and normal tissue prototypes is as high as 57%. This is a big handicap for the training of the classifiers, as we explain below.

To determine the influence of the density parameter in the performance of our system, we applied first a 30-fold cross validation scheme to train and test the system with the whole set of 5,052 prototypes. Next, a ROC analysis was performed over each of the 30 test results, calculating the area under curve (AUC) as a parameter to describe the performance over each test set. Finally, the mean value of the 30 AUCs was determined, to give a parameter that describes the overall performance of the system with those prototypes.

Table 2: Average number of prototypes of malignant (M) and benign (B) masses and normal tissue (N) divided into training, validation, and test sets, distributed by density value.

Density	Average distribution of prototypes by density value in the 30-fold cross validation study									
	Training			Validation			Test			
	M	B	N	M	B	N	M	B	N	Total
1	158.2	132.9	280.0	50.8	43.0	89.3	7.0	6.0	12.7	780.0
2	402.9	349.4	720.0	129.0	111.9	232.0	18.0	15.7	32.4	2012.0
3	236.3	238.9	559.6	76.2	76.3	178.5	10.5	10.8	25.0	1412.0
4	78.9	108.3	432.9	25.5	34.8	139.7	3.6	4.8	19.4	848.0
Overall	876.3	829.5	1192.5	281.5	266.0	639.5	39.1	37.3	89.5	5052.0

Table 3: This table shows the average results obtained over the different test subsets (considering all the prototypes, or only for those with a given density), as area under the ROC curve (AUC) for a confidence interval (CI) of 95%.

Mass-Normal tissue. Depending on density 30-fold cross validation test				
SVM		NN		Description
AUC	CI (95%)	AUC	CI (95%)	
0.944	[0.939, 0.949]	**0.943**	[0.938, 0.948]	Overall
0.964	[0.947, 0.980]	*0.965*	[0.948, 0.983]	cases with density 1
0.959	[0.951, 0.967]	0.961	[0.954, 0.969]	cases with density 2
0.927	[0.915, 0.939]	0.916	[0.902, 0.929]	cases with density 3
0.897	[0.872, 0.922]	***0.892***	[0.871, 0.914]	cases with density 4

This scheme was repeated later considering sets of prototypes containing only a given value of the density parameter, in order to compare the results. Those results are presented in Table 3. The overall results are presented in Figure 8 for both classifiers, and for cases with densities equal to 1 and 4 in Figure 9 for a NN classifier and Figure 10 for a SVM classifier.

As we expected, the best results were obtained for a density value equal to 1 (virtually fatty breasts with very little breast tissue, usually corresponding to old women), and the worst results for a density of 4 (very dense breasts, with much breast tissue, usually corresponding to young women). These results are consistent with other studies about the nature of cancer cases that are discarded by radiologists in a larger proportion [3–5].

Besides, it is important to remark that there are very different distributions of prototypes for the different values of density. While for a density of 1 the number of mass and normal tissue prototypes is almost the same (a 3% difference favorable to the number of mass prototypes), for a density of 4 the difference is very important (a 57% favorable to the number of normal tissue prototypes). This difference in the number of prototypes of each class introduces a statistical bias which could affect the training of the classifiers.

5. Conclusions

In this work, we have studied the influence of the BI-RADS density parameter assigned to a mammogram over the performance of our system. As a result, we have concluded that the performance is affected by that parameter, since the AUC of the ROC curves decreases from 0.965 to 0.892 (−7.56%) for NN classifiers and 0.964 to 0.897 (−6.95%) for SVM classifiers when we move from density 1 to density 4. However, taking into account that mammograms with density 4 are more difficult to analyze than those with density 1 (density 4 means very dense breasts with much breast tissue, so it is difficult to find masses, while density 1 means that very little breast tissue is present), and considering also the difficulties during training due to the different number of prototypes of both classes, we can conclude that our system is rather robust and performs very well even in the worst conditions.

Besides, it is important that the AUC for the global set of prototypes is only 2.28% and 2.07%, respectively, for NN and SVM classifiers, lower than the performance achieved for density 1, which is the most favourable case, so the performance of the system with the overall set is acceptable.

Finally, as the number of samples in the subsets of prototypes with densities equal to 2 and 3 is significantly higher than those in the subsets with densities equal to 1 and 4, we conclude that the variation of performance due to the BI-RADS density of our system is limited to about 7% in both cases.

On the other hand, it worth to remark the equality of performance obtained with the two types of classifiers tested.

Conflict of Interests

The authors declare that they have no conflict of interests.

Authors' Contribution

A. G.-Manso developed the preprocessing system (selection and acquisition of ROIs, and obtaining the ICA bases),

integrated the global system, conducted the experiments, obtaining and analysing the results, and drafted the paper. C. J. G.-Orellana developed the neural network classifier training algorithm, helped to adapt and adjust the hardware for the simulations (two Beowulf clusters with 45 and 48 nodes, resp.), together with the adaptation of the software to be run on the clusters. R. G.-Caballero developed the database associated with the experiments. H. M. G.-Velasco was responsible for the assembly and tuning of the clusters. M. M.-Macias developed the training algorithm for the support vector machine classifiers.

Acknowledgments

This work was supported in part by the "Junta de Extremadura" and FEDER through Projects PRI08A092, PDT09A036, GR10018, and PDT09A005.

References

[1] O. Peart, *Mammography and Breast Imaging. Just the Facts*, McGraw-Hill, New York, NY, USA, 1st edition, 2005.

[2] D. Kopans, *Breast Imaging*, Lippincott Williams & Wilkins, Baltimore, Md, USA, 2007.

[3] R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology*, vol. 184, no. 3, pp. 613–617, 1992.

[4] T. M. Kolb, J. Lichy, and J. H. Newhouse, "Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations," *Radiology*, vol. 225, no. 1, pp. 165–175, 2002.

[5] M. T. Mandelson, N. Oestreicher, P. L. Porter et al., "Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers," *Journal of the National Cancer Institute*, vol. 92, no. 13, pp. 1081–1087, 2000.

[6] R. F. Brem, J. W. Hoffmeister, J. A. Rapelyea et al., "Impact of breast density on computer-aided detection for breast cancer," *American Journal of Roentgenology*, vol. 184, no. 2, pp. 439–444, 2005.

[7] J. Freixenet, A. Oliver, R. Martí et al., "Eigendetection of masses considering false positive reduction and breast density information," *Medical Physics*, vol. 35, no. 5, pp. 1840–1853, 2008.

[8] A. Oliver, J. Freixenet, J. Martí et al., "A review of automatic mass detection and segmentation in mammographic images," *Medical Image Analysis*, vol. 14, no. 2, pp. 87–110, 2010.

[9] A. Horsch, A. Hapfelmeier, and M. Elter, "Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies," *International Journal of Computer Assisted Radiology and Surgery*, vol. 6, no. 6, pp. 749–767, 2011.

[10] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography," in *proceedings of the 5th International Workshop on Digital Mammography (IWDM '01)*, M. J. Yae, Ed., pp. 212–218, 2001.

[11] R. Campanini, D. Dongiovanni, E. Iampieri et al., "A novel featureless approach to mass detection in digital mammograms based on support vector machines," *Physics in Medicine and Biology*, vol. 49, no. 6, pp. 961–975, 2004.

[12] E. Angelini, R. Campanini, E. Iampieri, N. Lanconelli, M. Masotti, and M. Roffilli, "Testing the performances of different image representations for mass classification in digital mammograms," *International Journal of Modern Physics C*, vol. 17, no. 1, pp. 113–131, 2006.

[13] B. W. Hong and M. Brady, "A topographic representation for mammogram segmentation," *Lecture Notes in Computer Science*, vol. 2879, pp. 730–737, 2003.

[14] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision With the OpenCV Library*, O'Reilly Media, 1st edition, 2008.

[15] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis. Adaptive and Learning Systems for Signal Processing, Communications, and Control*, John Wiley & Sons, New York, NY, USA, 2001.

[16] M. Rolli, *Advanced machine learning techniques for digital mammography [Ph.D. thesis]*, Department of Computer Science, University of Bologna, Bologna, Italy, 2006.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.

[18] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, New York, NY, USA, 1999.

[19] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Statistics for Engineering and Information Science, Springer, New York, NY, USA, 2nd edition, 2000.

[20] R. M. Rangayyan, F. J. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs," *Journal of the Franklin Institute*, vol. 344, no. 3-4, pp. 312–348, 2007.

[21] I. Gonzalez-Carrasco, A. Garcia-Crespo, B. Ruiz-Mezcua, and J. L. Lopez-Cuadrado, "An optimization methodology for machine learning strategies and regression problems in ballistic impact scenarios," *Applied Intelligence*, vol. 36, no. 2, pp. 424–441, 2012.

[22] M. Riedmiller and H. Braun, "Direct adaptive method for faster backpropagation learning: the RPROP algorithm," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 586–591, April 1993.

[23] A. Zell, N. Mache, R. Huebner et al., "SNNS (Stuttgart Neural Network Simulator)," *Neural Network Simulation Environments*, vol. 254, pp. 165–186, 1994.

[24] C. C. Chang and C. J. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

[25] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[26] B. Ripley, "FastICA algorithms to perform ICA and projection pursuit," 2009, http://cran.r-project.org/web/packages/fastICA/fastICA.pdf.

[27] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 240182, 6 pages
http://dx.doi.org/10.1155/2013/240182

Research Article

An Intelligent System Approach for Asthma Prediction in Symptomatic Preschool Children

E. Chatzimichail,[1] E. Paraskakis,[2] M. Sitzimi,[2] and A. Rigas[1]

[1] *Department of Electrical and Computer Engineering, Democritus University of Thrace, 67100 Xanthi, Greece*
[2] *Department of Pediatrics, Democritus University of Thrace, 68100 Alexandroupolis, Greece*

Correspondence should be addressed to E. Chatzimichail; echatzim@ee.duth.gr

Received 26 October 2012; Accepted 21 February 2013

Academic Editor: Angel García-Crespo

Objectives. In this study a new method for asthma outcome prediction, which is based on Principal Component Analysis and Least Square Support Vector Machine Classifier, is presented. Most of the asthma cases appear during the first years of life. Thus, the early identification of young children being at high risk of developing persistent symptoms of the disease throughout childhood is an important public health priority. *Methods*. The proposed intelligent system consists of three stages. At the first stage, Principal Component Analysis is used for feature extraction and dimension reduction. At the second stage, the pattern classification is achieved by using Least Square Support Vector Machine Classifier. Finally, at the third stage the performance evaluation of the system is estimated by using classification accuracy and 10-fold cross-validation. *Results*. The proposed prediction system can be used in asthma outcome prediction with 95.54 % success as shown in the experimental results. *Conclusions*. This study indicates that the proposed system is a potentially useful decision support tool for predicting asthma outcome and that some risk factors enhance its predictive ability.

1. Introduction

Asthma is a chronic inflammatory disorder of the airways characterized by an obstruction of airflow, which may be completely or partially reversed with or without specific therapy [1]. Airway inflammation is the result of interactions between various cells, cellular elements, and cytokines. In susceptible individuals, airway inflammation may cause recurrent or persistent bronchospasm, with symptoms like wheezing, breathlessness, chest tightness, and cough, particularly at night or after exercise. Asthma is a disease with polymorphic phenotype affected by several environmental and genetic factors which both play a key role in the development and persistence of the disease [2, 3]. Among these factors family history of asthma, presence of atopic dermatitis or allergic rhinitis, wheezing episodes during childhood, maternal smoking during pregnancy, and several prenatal and environmental factors are included [4–7].

Most children who suffer from asthma develop their first symptoms before the 5th year of age [8]. However, it is difficult to discriminate asthma from other wheezing disorders of the childhood because the symptoms are similar. Thus, children with asthma may often be misdiagnosed as having a common cold, bronchiolitis, or pneumonia. For the diagnosis of asthma a detailed medical history and physical examination along with a lung function test is usually required. On the other hand, lung function test is hard to be performed in children younger than five years old.

In preventive medicine, the value of a test lies in its ability to identify those individuals who are at high risk of an illness and who therefore require intervention while excluding those who do not require such intervention. The accuracy of the risk classification is of particular relevance in the case of asthma disease. Early identification of patients at high risk for asthma disease progression may lead to better treatment opportunities and hopefully better disease outcomes in adulthood [9–13].

Several efforts have been made by different groups to discover a safe way of prediction of asthma outcome such as asthma index API or modified asthma index mAPI

in children younger than five years old [14, 15]. To the knowledge of the authors, this is the first study where machine learning techniques are used in the prediction of persistent asthma. However, Principal Component Analysis (PCA) has been used in several medical studies as for instance to evaluate the multivariate association between functional microvascular variables and clinical-laboratorial-anthropometrical measurements [16]. Moreover, in the study of [17], multivariate projection techniques have been utilized to reveal how inflammatory mediators demonstrate a distinct pattern of response to traumatic brain injury in humans. Finally, in [18], PCA was used for Gait Kinematics Data in Acute and Chronic Stroke Patients. Least Square Support Vector Machine (LSSVM) classifiers have been used with success for diagnosis of lung cancer [19] and in a hepatitis diagnosis system [20].

PCA provides a powerful method for exploring complex datasets with multiple variables and missing data points with relatively small numbers of observations [21]. LSSVM is a robust and reliable classifier system and has the ability to perform fast classification. For these reasons, those two techniques have been chosen for this study [22].

In this paper an intelligent system approach for asthma prediction outcome is presented. The system consists of three stages: (a) feature extraction and reduction through PCA, (b) pattern classification by using LSSVM classifier, and (c) the performance evaluation of the classifier by means of accuracy, sensitivity, specificity, and 10-fold cross validation. The paper is organized as follows. Section 2.1 presents the experimental dataset which has been used for this study. In Section 2.2 brief description of the PCA is shown, while in Section 2.3 the LSSVM classifier is introduced. In Section 2.4 the proposed prediction system is presented, while the results are shown in Section 3. The discussion and the final conclusions are described in Sections 4 and 5, respectively.

2. Methods

2.1. Clinical Data. Data from 148 patients from the Pediatric Department of the University Hospital of Alexandroupolis, Greece were collected during the period 2008–2010 and recorded. A group of 148 patients who received a diagnosis of asthma were studied prospectively from the 7th to 14th year of age. All patients with missing data were excluded from the present study, leaving a total of 112 patients.

A case history, including data on asthma, allergic diseases, and lifestyle factors was obtained by questionnaires. The participants (parents and their children) answered questions regarding asthmatic and allergic symptoms, wheezing episodes until the 5th year, pet keeping, family members, parental history, and some other useful information. The prognostic factors that were used in the questionnaire have been derived from previous studies [2–10]. A total of 46 prognostic factors have been considered and they are summarized in Table 1. For some of them a kind of encoding was required in order to be efficiently utilized for the current investigation. Their encoding is presented in Table 2.

Table 1: Prognostic factors.

Category	Prognostic factors
Demographic	age, sex, ethnicity[#], height, weight, waist's perimeter, residence[#]
Wheezing episodes	until 3rd year, between 3rd and 5th year
Symptoms	wheezing[*], cough[*], allergic rhinitis[*], runny nose[*], congestion[*], eczema[*], food allergy[*], pharmaceutical allergy[*], allergic conjunctivitis[*], dyspnea[*], seasonal symptoms[#]
Parental history	asthma[*]
House conditions	number of family members, pets[*], type of heating[#]
Pharmaceutical therapy	bronchodilators, corticosteroids inhaled[*], corticosteroids per os[*], antileukotriene[*], antihistamine[*]
Breathing tests	FEV_1%, $FEF_{25/75}$%
Tests	Ig E U/Ml
Allergens	*d. pteronyssinus*[#], *d. farinae*[#], olive[#], pellitory[#], graminaceae[#], pine[#], cypress[#], cat[#], dog[#], *alternaria*[#]
Neonatal period	pregnancy duration, breastfeeding duration[#], smoking during pregnancy[*]
Asthma	treatment[*]

[*] The encoding is binary: yes (1) or no (0).
[#] The encoding is shown in Table 2.
All other factors are numerical.

2.2. Principal Component Analysis for Feature Reduction. In the present study, the dimension of the input vector is large, while at the same time the components of the vectors are strongly correlated. It is, therefore, useful in this case to reduce the dimension of the input vectors. An effective procedure to perform this operation is to employ the PCA method. This technique has three effects: it orthogonalizes the components of the input vectors so that they are uncorrelated with each other, it sorts the resulting orthogonal components (principal components) so that those with the largest variation come first, and, finally, it eliminates those components that contribute the least to the variation in the data set [23].

According to the literature [24], the most common definition of PCA is that for a set of observed vectors $\{v_i\}$, $i \in \{1, 2, \ldots, N\}$, the q principal axes w_j, $j \in \{1, 2, \ldots, q\}$ are those orthonormal axes onto which the retained variance under projection is maximal. It can be shown that the vectors w_j are given by the q dominant eigenvectors (i.e., those with largest associated eigenvalues) of the covariance matrix

$$C = \sum_i \frac{(v_i - \overline{v})(v_i - \overline{v})^T}{N} \quad (1)$$

such that $Cw_j = \lambda_i w_j$, where $\overline{v}$ is the simple mean value and λ_i is a scalar, termed the eigenvalue corresponding to w_j.

The vector $u_i = W^T(v_i - \overline{v})$, where $W = (w_1, w_2, \ldots, w_q)$, is a q-dimensional reduced representation of the observed vector v_i [25].

Table 2: Encoding of some prognostic factors.

Prognostic factor	Coding					
Sex	0 (Male)	1 (Female)				
Residence	0 (Urban)	1 (Semiurban)	2 (Rural)			
Season of the symptoms	0 (None)	1 (Winter)	2 (Autumn)	3 (Spring)	4 (Summer)	5 (>2 Seasons)
Type of heating	0 (Central heating)	1 (Wood stove)	2 (Oil stove)	3 (Fireplace)	4 (Central heating + Fireplace)	
Pregnancy duration in weeks	0 (<37)	1 (37-38)	2 (>38)			
Allergens	0 (0)	1 (3.5–6 mm)	2 (>6 mm)			

2.3. Least Square Support Vector Machine Classifier. Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. The foundations of SVMs have been developed by Vapnik [26] and gained popularity due to many promising features. SVMs perform classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. The goal of SVM is to produce a model in the form of $f(x) = \omega^T x + b$ which predicts the target values of the test data given only the test data attributes. The training set $\{x_i, y_i\}_{i=1}^{l}$, where $x_i \in \Re^n$ is the input and $y_i \in \{-1, +1\}$ is the output, shows the class.

The Representer Theorem [27] states that the solution ω can always be written as a linear combination of the training data:

$$\omega = \sum_{j=1}^{N} a_j y_j x_j. \tag{2}$$

In that way, the SVM can be formulated to learn a linear classifier

$$f(x) = \sum_{i=1}^{N} a_i y_i K(x, x_i) + b, \tag{3}$$

by solving an optimization problem over a_i, where a_i are Lagrange, b is a real constant, and N is the size of the training data.

$K(x_i, x_j)$ is a nonlinear kernel function given by $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, where $\varphi(x)$ is the nonlinear map from original space to the high dimensional space.

The SVM classifiers solve the following quadratic programming problem:

$$\min \frac{1}{2}\omega^T \omega + C \sum_{i=1}^{N} \xi_i \tag{4}$$

subject to $y_i(\omega^T \phi(x_i) + b) = 1 - \xi_i$, $\xi_i \geq 0$, $i = 1, \ldots, N$, ξ_i represents the degree of misclassification of the data x_i and C is the penalty parameter of the error term [28].

In this paper the least squares version of SVM is used, whose main advantage is that it is computationally more efficient than the standard SVM method. In this case the training process requires the solution of a linear equation set instead of the quadratic programming problem involved by the standard SVM. The LSSVM method—when Radial Basis Function (RBF) kernels are used—requires only two parameters (C and σ), while the time consumed by the training method is reduced, by replacing the quadratic optimization problem with a simple linear equation set [29]. In LSSVMs, an equality constraint-based formulation is made within the context of ridge regression as follows:

$$\min \frac{1}{2}\omega^T \omega + C \sum_{i=1}^{N} e^2{}_i \tag{5}$$

subject to $y_i(\omega^T \phi(x_i) + b) = 1 - e_i$, $i = 1, \ldots, N$.

2.4. The Intelligent PCA-LSSVM Prediction System. The asthma prediction system which is presented in this study consists of three stages: (i) the feature extraction and dimension reduction through PCA, (ii) the pattern classification by employing LSSVM classifier, and (iii) the performance evaluation by using classification accuracy, sensitivity, specificity, and 10-fold cross-validation. The flowchart of the intelligent system for asthma prediction is illustrated in Figure 1. The implementation steps of the algorithm follow a certain sequence. First of all, the patient's data were collected and prepared in an electronic form suitable for further processing. After this step, all the parameters (where it is necessary) were encoded and the outputs were assigned either with label 1 (asthma persistence) or 0 (no asthma persistence). At last, the dimension of the dataset which had 46 features was reduced to 18 features using the PCA method.

In the classification stage of PCA-LSSVM intelligent prediction system, the reduced features obtained from the first stage were fed to the LSSVM classifier. LSSVM classifiers parameters, which are σ (the width of RBF kernel) and margin-losses trade-off C, affect the prediction performance. The best combination of C and σ was selected by the grid search with growing sequences of C (1–1000 with a step equals 10) and σ (1–100 with a step equals 1). Each combination of parameter choices was checked using 10-fold cross-validation. At first, the 112 patients were divided into 10 almost equal subgroups. One of the 10 subgroups has been used as the evaluation data and the rest as the learning data for the classification. The evaluation data were changed 10 times, so that each group was investigated once as evaluation

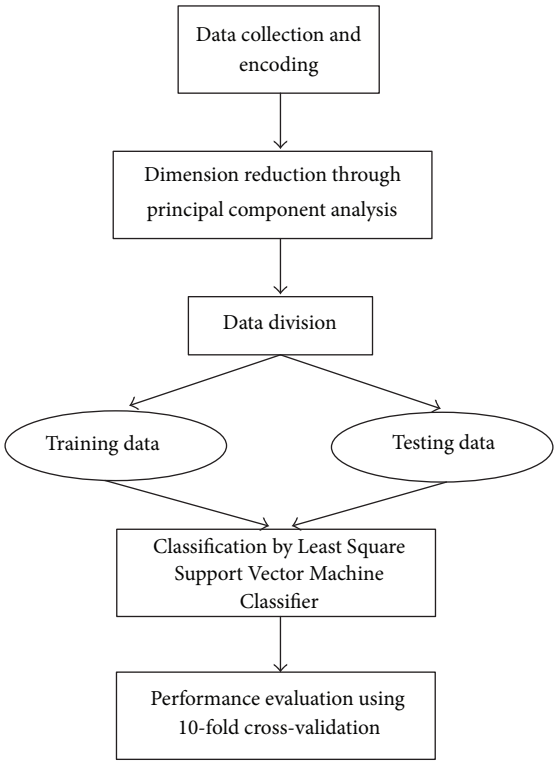


Figure 1: Flowchart diagram of the proposed intelligent system for asthma prediction.

data. The average value of all obtained accuracies of the evaluation data was considered as the estimation ability of the model. The parameters with best cross-validation accuracy were picked.

3. Results

The experimental results are presented in terms of accuracy, sensitivity, and specificity as shown in Table 3. The prediction is considered to be true positive (TP) if the patient has asthma and it is correctly predicted as asthmatic. On the contrary, if the asthmatic patient is incorrectly predicted as nonasthmatic, the prediction is assigned as false negative (FN) [30]. False positive (FP) and true negative (TN) predictions can be determined in the same way. Looking into the training model, there are 68 positive data (presence of asthma according to the physicians) and 44 negative (absence of asthma according to the physicians). The sensitivity, specificity, and accuracy have been estimated using the following equations:

$$\text{Sensitivity} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \times 100,$$

$$\text{Specificity} = \frac{N_{\text{TN}}}{N_{\text{TN}} + N_{\text{FP}}} \times 100, \quad (6)$$

$$\text{Accuracy} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}} \times 100,$$

Table 3: Accuracy, sensitivity, and specificity percent values for 10 combinations of C and σ.

Selection of σ parameter	Selection of C parameter	Accuracy	Sensitivity	Specificity
30	10	93.75	97.73	91.18
28	100	94.64	97.73	92.65
62	100	93.75	97.73	91.18
7	**10**	**95.54**	**95.45**	**95.59**
15	10	94.64	97.73	92.65
72	1000	93.75	97.73	91.18
49	100	93.75	97.73	91.18
58	10	90.18	95.45	86.76
25	10	93.75	97.73	91.18
50	1000	94.64	97.73	92.65

where N_{TP}, N_{TN}, N_{FP}, N_{FN} are the number of TP, TN, FP, FN, respectively [31]. Sensitivity and specificity are statistical measures of the performance of a binary Classification test. Sensitivity measures the percentage of positive (asthmatic) people that have been correctly identified as having asthma. Specificity measures the percentage of negative (not asthmatic) people which have been correctly identified as not having asthma. The accuracy is the degree of how close the predicted values are to the actual ones.

In Table 3, the best-performed 10 combinations of C and σ values and the correct asthma prediction rates are presented. As it can be seen from these results, the value having the highest prediction accuracy for the proposed asthma prediction intelligent method was found to be 95.54%, for the case where $\sigma = 7$ and $C = 10$.

4. Discussion

The predictive accuracy of the proposed system is not easily comparable with that of other studies because of differences in study design and objectives. To the authors' knowledge a limited number of studies have been published on asthma prediction in children at the age when the symptoms are observed. In the study of Caudri et al. [32] the asthma prediction was based on eight clinical parameters, considering children from 7 to 8 years of age. These eight parameters were male sex, postterm delivery, parental education and inhaled medication, wheezing frequency, wheeze/dyspnea apart from colds, respiratory infections, and eczema. In 72% of the cases, the model accurately discriminated the asthmatic and the nonasthmatic children. Clough et al. [33] have developed models to examine the potential risk factors for wheeze that persists for at least 12 months after presentation in a group of young children, each with at least one atopic parent, with early-life wheezing. This paper has shown that increased age at presentation, personal atopy and raised soluble IL-2R are all associated with increased risk. Castro-Rodríguez et al. [34] developed two clinical indices at 3 years of age for the prediction of asthma in school age. It was shown that 59% of children with a positive loose index and 76% of those with a positive stringent index had active asthma

in at least one survey during the school years. Their indices include characteristics of wheezing during the first 3 years of life, parental asthma or eczema, wheezing without colds, eosinophilia, or allergic rhinitis. Finally, in the study of Devulapalli et al. [35] the number of hospital admissions for obstructive airways disease within the first 2 years of life has been included in the predictive model, giving positive predictive values and negative predictive value of 55% and 92%, respectively.

Based on the comparison, which has been already shown above, it seems that there are valuable studies published on asthma prediction. However, these prediction methods are not able to achieve substantially high predictive accuracies. It is, therefore, meaningful to utilize computational intelligence methods in order to overcome such problems. Such an example has been shown in this paper. The proposed method for asthma prediction up to the age of 5 might predict the asthma with an accuracy exceeding 95%. However future studies should be performed to further evaluate our proposed method in clinical practice. Moreover, regardless of the prediction outcome using the presented algorithm, an evaluation of the results in cooperation with medical doctors who are asthma specialists must be performed in order to decide if either the patient needs treatment or not.

5. Conclusions

In this paper, a new intelligent system based on the Principal Component Analysis and Least Square Support Vector Machine classifier for asthma prediction has been proposed. The used parameter vector had a significantly high dimensionality and it was, therefore, necessary to be reduced in order to achieve as low computational cost as possible on the one hand, while on the other hand to minimize the complexity of the system. Due to the fact that asthma is a serious health condition, the various models, which have been used to detect it, must have high accuracy so that patients with asthma are not overlooked.

The experimental results show that the proposed method can predict 95.54% of patients with asthma. To conclude, the proposed system can give a significant contribution and be a useful tool in clinical practice for the physicians in order to overcome many of the therapeutic dilemmas.

Conflict of Interests

The authors report no conflict of interests.

References

[1] W. Eder, M. J. Ege, and E. von Mutius, "The asthma epidemic," *The New England Journal of Medicine*, vol. 355, no. 21, pp. 2226–2235, 2006.

[2] M. B. Bracken, K. Belanger, W. O. Cookson, E. Triche, D. C. Christiani, and B. P. Leaderer, "Genetic and perinatal risk factors for asthma onset and severity: a review and theoretical analysis," *Epidemiologic Reviews*, vol. 24, no. 2, pp. 176–189, 2002.

[3] N. E. Lange, S. L. Rifas-Shiman, C. A. Camargo, D. R. Gold, M. W. Gillman, and A. A. Litonjua, "Maternal dietary pattern during pregnancy is not associated with recurrent wheeze in children," *Journal of Allergy and Clinical Immunology*, vol. 126, no. 2, pp. 250–255, 2010.

[4] C. Porsbjerg, M. L. von Linstow, C. S. Ulrik, S. Nepper-Christensen, and V. Backer, "Risk factors for onset of asthma: a 12-year prospective follow-up study," *Chest*, vol. 129, no. 2, pp. 309–316, 2006.

[5] C. Tolomeo, C. Savrin, M. Heinzer, and A. Bazzy-Asaad, "Predictors of asthma-related pediatric emergency department visits and hospitalizations," *The Journal of Asthma*, vol. 46, no. 8, pp. 829–834, 2009.

[6] H. S. Zahran, C. J. Person, C. Bailey, and J. E. Moorman, "Predictors of asthma self-management education among children and adults—2006-2007 behavioral risk factor surveillance system asthma call-back survey," *The Journal of Asthma*, vol. 49, no. 1, pp. 98–106, 2012.

[7] N. N. Hansel, E. C. Matsui, R. Rusher et al., "Predicting future asthma morbidity in preschool inner-city children," *The Journal of Asthma*, vol. 48, no. 8, pp. 797–803, 2011.

[8] G. Nagel, G. Weinmayr, A. Kleiner et al., "Effect of diet on asthma and allergic sensitisation in the International Study on Allergies and Asthma in Childhood (ISAAC) Phase Two," *Thorax*, vol. 65, no. 6, pp. 516–522, 2010.

[9] F. D. Martinez, A. L. Wright, L. M. Taussig, C. J. Holberg, M. Halonen, and W. J. Morgan, "Asthma and wheezing in the first six years of life," *The New England Journal of Medicine*, vol. 332, no. 3, pp. 133–138, 1995.

[10] B. G. Toelle, W. Xuan, J. K. Peat, and G. B. Marks, "Childhood factors that predict asthma in young adulthood," *European Respiratory Journal*, vol. 23, no. 1, pp. 66–70, 2004.

[11] W. Balemans, C. van der Ent, A. Schilder, E. Sanders, G. Zielhuis, and M. Rovers, "Prediction of asthma in young adults using childhood characteristics: development of a prediction rule," *Journal of Clinical Epidemiology*, vol. 59, no. 11, pp. 1207–1212, 2006.

[12] F. Bacopoulou, A. Veltsista, I. Vassi et al., "Can we be optimistic about asthma in childhood? A Greek cohort study," *The Journal of Asthma*, vol. 46, no. 2, pp. 171–174, 2009.

[13] K. Porpodis, D. Papakosta, K. Manika et al., "Long-term prognosis of asthma is good—a 12-year follow-up study. Influence of treatment," *The Journal of Asthma*, vol. 46, no. 6, pp. 625–631, 2009.

[14] J. A. Castro-Rodriguez, "The asthma predictive index: a very useful tool for predicting asthma in young children," *Journal of Allergy and Clinical Immunology*, vol. 126, no. 2, pp. 212–216, 2010.

[15] J. A. Castro-Rodriguez, S. L. Cifuentes, and C. E. Rodríguez-Martínez, "The asthma predictive index remains a useful tool to predict asthma in young children with recurrent wheeze in clinical practice," *Journal of Allergy and Clinical Immunology*, vol. 127, no. 4, pp. 1082–1083, 2011.

[16] D. G. Panazzolo, F. L. Sicuro, R. Clapauch, P. A. Maranhão, E. Bouskela, and L. G. Kraemer-Aguiar, "Obesity, metabolic syndrome, impaired fasting glucose, and microvascular dysfunction: a principal component analysis approach," *BMC Cardiovascular Disorders*, vol. 12, article 102, 2012.

[17] A. Helmy, C. A. Antoniades, M. R. Guilfoyle, K. L. Carpenter, and P. J. Hutchinson, "Principal component analysis of the cytokine and chemokine response to human traumatic brain injury," *PLoS One*, vol. 7, Article ID e39677, 2012.

[18] I. Milovanović and D. B. Popović, "Principal component analysis of gait kinematics data in acute and chronic stroke patients," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 649743, 8 pages, 2012.

[19] E. Avci, "A new expert system for diagnosis of lung cancer: GDA-LS_SVM," *Journal of Medical Systems*, vol. 36, no. 3, pp. 2005–2009, 2012.

[20] D. Çalişir and E. Dogantekin, "A new intelligent hepatitis diagnosis system: PCA-LSSVM," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10705–10708, 2011.

[21] L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold, *Multi- and Megavariate Data Analysis: Principles and Applications*, Umetrics AB, 2001.

[22] H. Jiang and W. Ching, "Correlation Kernels for support vector machines classification with applications in cancer data," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 205025, 7 pages, 2012.

[23] J. R. G. Lansangan and E. B. Barrios, "Principal components analysis of nonstationary time series data," *Statistics and Computing*, vol. 19, no. 2, pp. 173–187, 2009.

[24] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[25] A. Widodo and B. S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Systems with Applications*, vol. 33, no. 1, pp. 241–250, 2007.

[26] V. Vapnik, "The support vector method," in *Proceeding of the 7th International Conference on Artificial Neural Networks (ICANN '97)*, pp. 263–271, 1997.

[27] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, no. 1, pp. 82–95, 1971.

[28] M. Slawski, "The structured elastic net for quantile regression and support vector classification," *Statistics and Computing*, vol. 22, no. 1, pp. 153–168, 2012.

[29] M. Sabzekar, H. S. Yazdi, and M. Naghibzadeh, "Relaxed constraints support vector machine," *Expert Systems*, vol. 29, no. 5, pp. 506–525, 2012.

[30] E. A. Chatzimichail, A. G. Rigas, and E. N. Paraskakis, "An artificial intelligence technique for the prediction of persistent asthma in children," in *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB '10)*, pp. 1–4, November 2010.

[31] E. Chatzimichail, A. Rigas, E. Paraskakis, and A. Chatzimichail, "Diagnosis of asthma severity using artificial neural networks," in *Proceedings of the 12th Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON '10)*, vol. 29, pp. 600–603, May 2010.

[32] D. Caudri, A. Wijga, C. M. Chipper et al., "Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age," *The Journal of Allergy and Clinical Immunology*, vol. 124, no. 5, pp. 903–910, 2009.

[33] J. B. Clough, K. A. Keeping, L. C. Edwards, W. M. Freeman, J. A. Warner, and J. O. Warner, "Can we predict which wheezy infants will continue to wheeze?" *American Journal of Respiratory and Critical Care Medicine*, vol. 160, no. 5, pp. 1473–1480, 1999.

[34] J. A. Castro-Rodríguez, C. J. Holberg, A. L. Wright, and F. D. Martinez, "A clinical index to define risk of asthma in young children with recurrent wheezing," *American Journal of Respiratory and Critical Care Medicine*, vol. 162, no. 4, pp. 1403–1406, 2000.

[35] C. S. Devulapalli, K. C. L. Carlsen, G. Håland et al., "Severity of obstructive airways disease by age 2 years predicts asthma at 10 years of age," *Thorax*, vol. 63, no. 1, pp. 8–13, 2008.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 317803, 11 pages
http://dx.doi.org/10.1155/2013/317803

Research Article

Ontology-Oriented Diagnostic System for Traditional Chinese Medicine Based on Relation Refinement

Peiqin Gu,[1] Huajun Chen,[1] and Tong Yu[2]

[1] *College of Computer Science, Zhejiang University, Hangzhou 310027, China*
[2] *China Academy of Chinese Medical Sciences, Beijing 100700, China*

Correspondence should be addressed to Huajun Chen; huajunsir@zju.edu.cn

Received 26 October 2012; Revised 31 December 2012; Accepted 2 January 2013

Academic Editor: Alejandro Rodríguez González

Although Chinese medicine treatments have become popular recently, the complicated Chinese medical knowledge has made it difficult to be applied in computer-aided diagnostics. The ability to model and use the knowledge becomes an important issue. In this paper, we define the diagnosis in Traditional Chinese Medicine (TCM) as discovering the fuzzy relations between symptoms and syndromes. An Ontology-oriented Diagnosis System (ODS) is created to address the knowledge-based diagnosis based on a well-defined ontology of syndromes. The ontology transforms the implicit relationships among syndromes into a machine-interpretable model. The clinical data used for feature selection is collected from a national TCM research institute in China, which serves as a training source for syndrome differentiation. The ODS analyzes the clinical cases to obtain a statistical mapping relation between each syndrome and associated symptom set, before rechecking the completeness of related symptoms via ontology refinement. Our diagnostic system provides an online web interface to interact with users, so that users can perform self-diagnosis. We tested 12 common clinical cases on the diagnosis system, and it turned out that, given the agree metric, the system achieved better diagnostic accuracy compared to nonontology method—92% of the results fit perfectly with the experts' expectations.

1. Introduction

Applying mathematical models and information technologies to medical intelligence has long been a hot spot in the academic research domains and real-life health care applications. Plenty of the efforts in this field allow researchers and medical practitioners to identify required information more efficiently, discover new substances or relationships, and integrate different sources of information more easily.

Traditional Chinese Medicine, known as TCM, that is an ancient and unique branch of medical science, which probably covers broad range of practices such as herb medicine, acupuncture, attracts much attention on how to defeat the inconsistency of therapeutic patterns and vagueness of medical terms in TCM in order to improve the user experience of TCM diagnosis.

The basic theories of TCM are based on the ancient philosophy of holistic understanding of the universe and the human body, which are commonly associated with the flow of Qi and the balancing of complementary opposites, such as Yin and Yang, or five elements. Based on the interacting forces in the human body, the forms of many pathological conditions in TCM usually differ from that of identifiable diseases in terminology of the Western medicine. "*Syndrome*" (also called pattern) refers to a pattern of disharmony or functional disturbance within the functional entities that the TCM model of the body is composed of. A "*disease*" in TCM refers to disease entity or disease category, focusing on the macroscopic classification of specific manifestations. In TCM diagnosis, the therapy is determined mainly according to the pattern rather than the disease. Two patients with the same disease but different patterns will receive different therapy, vice versa patients with similar patterns might receive similar therapy even if their diseases are different. Hence, the concept of "syndrome differentiation" is of great importance in deciding the diagnosis for patients. But the difficulty of differentiating syndromes is rooted in the unclear definition of the range of syndromes, which means the determination of specific syndromes, the similarity among syndrome patterns,

and these information are usually hidden in the literature texts or the doctors' experience.

In this paper, we propose a simple but useful diagnostic system based on a domain ontology proposed by professional TCM experts, which is supposed to formalize the hidden knowledge and be imported into the diagnostic process. Similar to other TCM diagnostic systems, this framework basically includes two important components: *clinical database of patient data* and *knowledge base of patterns or classifiers*. 10,000 high-quality cases were collected, involving 5 types of diseases, 732 types of syndromes, and 3,519 symptoms, from a big clinical database of China Academy of Chinese Medicine Sciences (CATCM), a national research institute in Beijing. In the feature extraction phase, symptoms related to the same syndrome or same disease from different clinical cases are calculated by an adaptive intersection algorithm to generate a minimum symptom set. The goal of this step is in fact feature selection in a reductive way, in which some key symptoms are the features in order to identify a syndrome or a disease. However, there are still three problems to solve: *concept subsumption*, *synonym*, *inclusion* relations among syndromes. The *Ontology-oriented Diagnostic System (ODS)* address these problems by applying *ontology representation* as one of the techniques of semantic technologies, to systematize the important but usually hidden knowledge of syndrome relationships. It captures all the subtle relationships among syndromes as a terminology base and stores it in an ontology file. The benefit of building specific ontology is to enable the standardized terminology description of scientific domain and smooth data interchange on the web. In ODS, the ontology of syndromes is used to rectify the initial diagnosis results derived from the predetermination of syndrome-symptom, disease-symptom relations by rechecking the associated syndromes, also called a *relation refinement* process. A Web-based GUI is especially designed for this system to allow user-friendly instant diagnosis without human intervention. To evaluate the usefulness of the system, we invited TCM experts to validate the accuracy of our system by scoring each diagnostic result, and the feedback turned out promising.

The major works in this paper are the following.

(i) Use clinical data as a feature selection source, and propose a new computation model for initial decision making of TCM diagnosis. The aim is to find the most closely related set of symptoms for each occurred syndrome or disease entity, which can be regarded as a statistical mapping table.

(ii) With the broad use of biomedical ontologies, a simple ontology is designed for describing the interrelationships between syndromes, as an attempt to define the range and overlap of syndromes. The inheritance and subsumption will be used to distinguish between them.

(iii) Based on the expertise-dependent characteristic of TCM diagnosis, an agree metric is proposed to evaluate the efficiency of our diagnostic system, which is actually ranking of the diagnostic results. The scores suggested that ODS shows certain potential in making diagnostic decisions.

The rest of the paper is organized as follows. Section 2 outlines related research in the area. Section 3 gives a clear formalization of the medical diagnosis problem. Section 4 gives a detailed introduction of the kinds of relationships we built in the syndrome ontology. In Section 5, we demonstrate the minimum set extraction algorithm applied on the clinical database. In Section 6, after constructing mapping relations between the syndromes and symptoms based on the clinical records, relation refinement based on ontology hierarchy will be applied on the initial diagnostic result to finalize it. Section 7 gives a demonstration of a web-based diagnostic platform and evaluates the accuracy, performance, and scalability of the diagnostic capability of the system. Finally, we conclude and discuss the future work in Section 8.

2. Related Work

In the domain of medical diagnosis systems, a multitude of approaches exist, including various algorithmic techniques for intelligent automatic diagnosis that has been used to solve practical diagnostic problems in TCM, as well as in Western medicine, such as Bayesian networks, ontology engineering, and rule-based inference.

The design of automatic TCM diagnostic systems usually has to deal with the unstructured clinical information and vague description of domain knowledge. Bayesian classifiers such as Naïve Bayesian classifier are the most popular methods that have been applied in TCM diagnostics to the learning and classification of disease and syndrome based on symptoms. Moreover, Bayesian network algorithms are used to construct knowledge bases from clinical data [1]. A self-learning system for diagnosis in TCM [2] constructed a knowledge base by applying an improved hybrid Bayesian network learning algorithm with data mining techniques. By performing Naïve-Bayes classifiers with a score-based strategy for feature selection and a method for mining constrained association rules, the diagnostic results of the system turned out encouraging.

In recent years, to facilitate biomedical research among communities, various ontologies and knowledge bases have been developed [3], for example the *Gene Ontology (GO)*, *UMLS* [4], and *SNOMED RT* [5], to provide the capability of knowledge management, knowledge sharing. Compared to Bayesian methods, the semantic web technologies benefit the biomedical domain by providing a conceptualisation of the domain with the means of specifying the concepts and the relationships into a standardized form. When being extended to medical domains, semantic technologies can reveal machine-readable latent relationships within information where the homogeneity of terminology is particularly critical. Knowledge-based systems for medical diagnosis such as [6] build a formal ontology to manage medical information and suggest that the use of semantic technologies could improve the accuracy of medical diagnosis. ODDIN, an ontology-driven medical diagnosis system [7], uses a number of knowledge-based technologies to solve the problems of ambiguity which is very common in medical diagnosis, such as ontologies representing specific structure information and

probabilistic statistical refinements. The semantic ontology support in ODDIN benefits the diagnostics with the reasoning and inference capabilities offered by ontologies.

In the field of TCM, the emergence of knowledge-based diagnostic systems might improve the understanding of medical diagnostic process which is based on philosophical theories. A knowledge-based Chinese Medical Diagnostic System (CMDS) [8] extracted an integrated medical ontology with described domain knowledge and provided a standardized understanding of digestive disease diagnosis. By constructing a hierarchical ontology and association rules as classifiers, the system facilitated the acquisition, verification, and maintenance of knowledge by both human and machines. This prototype of CMDS can diagnose about 50 types of diseases by using over 500 rules and 600 images for various diseases. For those who is looking for more work on extracting ontology statements to build a robust specialized knowledge base for TCM, [9] is a good reference work that have been done by Zhejiang University, which includes more than 10,000 classes and about 80,000 instances.

According to a recent survey on computational methods for TCM [10], although there have been several TCM approaches employing different ontologies, standardized ontology seems absolutely necessary. In order to get richer and more adequate diagnostic results in TCM diagnostic systems, information from domain ontologies should be well used.

3. Problem Formulation

A medical diagnosis problem can be formed as a probabilistic relation $R \subset S \times D$ between a group of symptoms S and a diagnosis D. Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of symptoms to be queried, in which m is the cardinality of S, namely, the number of symptoms. Each $s_i \in S$, $1 \leq i \leq m$ is a professional description of a pathological symptom, such as *cough* and *fever*. The diagnostic result $D = C \oplus Q$ is an integrated answer of some disease and at least one syndrome, meaning that the patient catches a disease named C with a collection of syndromes Q.

The difficulties in determining this relation R between S and D lie on the obscure descriptions of each clinic case and uncertain (expertise-dependent) differentiation of syndromes. As an outsider of medical science, we are suggested by the medical experts that a relation table P should be defined in advance, in which there are two kinds of relations R_1 and R_2. R_1 represents the correspondence between a specific *disease* and a set of *symptoms* and R_2 represents the correspondence between a specific *syndrome* between a set of *symptoms*, and a table of these correspondence relations form a feature selection source for medical diagnosis. For example, we could obtain a minimum associated set of symptoms for each disease or syndrome such as $\{Heat \rightarrow (heat, dizzy)\}$ by analyzing prior clinical data $\{Heat \rightarrow (heat, dizzy, redface)\}$,$\{Heat \rightarrow (heat, dizzy, sore\ throat)\}$, which is formatted as (disease C or syndrome Q, related symptoms S). The process of clustering and filtering data is a variant of set intersection problem. We put our considerations on the *performance* and *efficiency* of the algorithm.

In order to rectify the data biases of clinical data and facilitate knowledge-based diagnosis, we define a knowledge base B which takes the role of the conceptual model. In the procedure of automatic diagnosis, the user-queried symptoms S are compared with the relation R generated from the preprocessed relation table P and the knowledge base B. By calculating the similarity, the diagnostic result D will be given as the combined answer of a disease C and a group of syndromes Q.

The process of Chinese medical diagnosis can be concluded as an expertise-dependent case search based on observed pathological patterns, where expertise is consisted of practical experiences and theoretical principles. In this paper, we integrate both practical experiences and theoretical principles together to construct a complete computation model for reliable medical diagnosis.

4. Ontology Modelling for TCM Syndromes

In TCM, syndrome refers to the association of several clinically recognizable features, usually identified by a group of symptoms that collectively indicate or characterize a disease, psychological disorder or other abnormal condition. More and more medical researchers recognize that the combination of disease diagnosis in biomedicine and pattern differentiation in TCM is essential for the clinical practice, and it has been a common practice model in China, since it will produce better clinical effects. Syndrome differentiation will help improve the clinical efficacy in clinical practice since it further specifies the indication with TCM classification. Syndrome differentiation is mainly based on symptoms, including tong appearance, pulse palpation, and patient's mental state, which aims to classify the symptoms into specific groups. However, the syndromes found in TCM are not totally independent, instead there are certain connections among them, for example, subsumption, equivalence, and disjointness.

Semantic web technologies are currently widely used in the biomedical domain [11], capturing, structuring, retaining, and reusing information to develop an understanding of the whole system (e.g., genome, pathway), and subsequently to convey this information meaningfully to other information systems. One of the methods to represent biomedical knowledge is to build a comprehensive terminology ontology, sometimes specific to the applications at first, but may be easily adapt to further integration. The basic idea of an ontology is to construct abstract knowledge by resource triples $\langle s, p, o \rangle$, which states that a relation denoted by p exists between subject s and object o. For example, we could declare a statement that $\langle Insomnia, causes, easy\ to\ wake \rangle$, in which *Insomnia* is the subject disease, *easy to wake* is a pathological symptom *caused* by the disease. The representation languages used to build ontology is proposed by W3C consortium, such as Resource Description Framework (RDF) [12] and Web Ontology Language (OWL) [13].

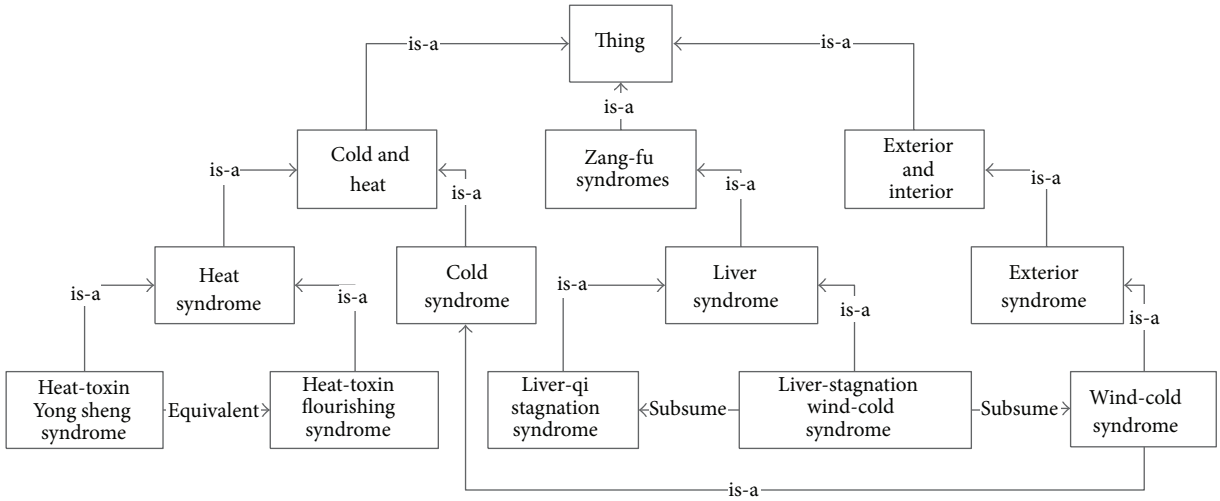


Figure 1: A portion of the hierarchy of syndrome ontology model.

An ontology of syndromes called *SynOnt* is designed to structure the interrelationships among TCM syndromes, which should also be available for further extension and integration (file published at https://www.github.com/astergu/TCM_Link/blob/master/syndrome.owl/). Generally, an ontology of high quality is developed from curated knowledge maintained by domain experts [14]. In order to extract a set of concise statements as a reliable knowledge base, one research scientist in CATCM was invited to build the ontology manually in 2009 for a month. We did a systematic bibliographic analysis on three widely acknowledged teaching materials "Basic Theories of Chinese Medicine" [15], "Therapeutic Science in Chinese Medicine" [16], and "Inner Medicine of Chinese Medicine" [17]. The ontology was developed using the open source ontology editor Protégé, by first creating top categorical concepts and corresponding subconcepts, then for related syndromes, connect them with property relations. The standard naming of the syndromes and the relations between them were fully discussed and modelled as semantic classes and properties. As shown in Figure 1, *SynOnt* ontology deals with three major problems related to syndrome differentiation, `equivalence` (equivalent), `inheritance` (is-a), and `subsumption` (subsume).

4.1. Equivalent Syndromes. There are some circumstances that exist multiple names for one identical thing, especially happen a lot in medicine, for example, the plant *Rheum palmatum* is also called *Fire ginseng*. This kind of knowledge appears as common sense for domain experts, but rather difficult for computers to know and understand. Identifying obvious similar references help merge different sources of knowledge together and reduce the cost of data mining and knowledge discovery.

The built-in OWL property *owl:equivalentClass* links a class description to another class description, which indicates that the two URI references actually refer to the same concept; namely, they have the same class extension. The *owl:equivalentClass* statements are often used in definition mapping between ontologies. For an individual syndrome class such as *"deficient cold syndrome,"* we could state that the following three URI references actually refer to the same kind of syndrome:

```
<owl:Class rdf:about="Deficient cold syndrome">
<owl:equivalentClass rdf:resource="Cold
syndrome">
<owl:equivalentClass rdf:resource="Yang
deficient syndrome">
</owl:Class>
```

By that mean, we say *deficient cold syndrome* is the same as both *yang deficiecy syndrome* and *cold syndrome*, which can be easily imported to other data sources with inference tools.

4.2. Syndrome Inheritance. The inheritance relationship between syndrome classes is defined in the categorical hierarchy of syndromes in the form of RDF triple $\langle C_1 rdfs : subClassOf C_2 \rangle$. This relationship shows the affiliation relation between C_1 and C_2. If the class description C_1 is defined as a subclass of class description C_2, then the set of individuals in the class extension of C_1 should be a subset of the set of individuals in the class extension of C_2. This class axiom declares a subclass relation between two OWL classes that are described through their names.

The inheritance relationships between syndromes shows the categorical information in TCM diagnostics. Generally, the type of syndromes are divided into four major categories: *heat and cold*, *zang-fu*, *deficiency and excess* and *exterior and interior*, whose classification reflects the pathological deviation of syndromes. For a syndrome associated with a set of symptoms, we regard the set of symptoms as the extension of the specific syndrome. According to the diagnostic principle, the symptoms related to the subclass syndrome inherits the symptoms related to the superclass syndrome.

4.3. Syndrome Subsumption. Based on the understanding of TCM therapeutic theories, we found that different resources

of syndromes could not only be duplicated by name, but also associated with each other by relational algebra. For example, the syndrome *deficiency of lung-yin and kidney-yin* actually means the abnormal conditions of yin deficiency happen simultaneously both in lung and kidney, not varying much from the composition of *lung-yin deficiency* and *kidney-yin deficiency*.

We define a property relation *tcm:subsume* which extends the meaning of built-in property *owl:unionOf*. The *owl:unionOf* property links a class to a list of class descriptions. An *owl:unionOf* statement describes an anonymous class for which the class extension contains those individuals that occur in at least one of the class extensions of the class descriptions in the list. However, the inclusion relation *tcm:subsume* defined by us assigns the union of symptoms associated with a collection of individual syndromes, with a "or" relation.

```
<owl:Class rdf:about="Qi stagnation and blood
stasis">
<owl:equivalentClass>
<owl:Restriction>
<owl:onProperty rdf:resource="subsume"/>
<owl:someValuesFrom rdf:resource="Qi
stagnation syndrome"/>
</owl:Restriction>
</owl:equivalentClass>
<owl:equivalentClass>
<owl:Restriction>
<owl:onProperty rdf:resource="subsume">
<owl:someValuesFrom rdf:resource="Blood stasis
syndrome"> </owl:Restriction>
</owl:equivalentClass>
<rdfs:subClassOf rdf:resource="Excess
syndrome">
</owl:Class>
```

If it appears that *a tcm:subsume b* in user-friendly Manchester syntax, then it means syndrome *a* captures all the symptoms when syndrome *b* occurs, but not vice versa. If *a tcm:subsume b* and *a tcm:subsume c*, then *a* links to all the symptoms related to both *b* and *c*.

5. Syndrome-Symptom Relation Extraction from Clinical Data

We propose a feature set extraction method based on a revised adaptive set intersection algorithm to preprocess the rows of clinical data. In fact, in the three-column design of the clinical database, C stands for the column of `disease`, Q for `syndrome` and P for `symptoms`. In general, disease in TCM is perceived as a category to distinguish the disharmony (or imbalance) in the functions or interactions of yin, yang, qi, xue, zang-fu, and meridians and the interaction between the human body and the environment. In clinical practice, the identified pattern (syndrome) usually involves differentiation of occurred symptoms. Each row represents a clinical statement that once a patient caught all the symptoms as listed in P, then he or she was diagnosed to have the disease named C and the syndrome named Q. These clinical data are thought to be trustworthy and of high quality; since they are extracted from another journal paper database of CATCM, the data of which are collected from top TCM journals, such as "*Zhong Hua Zhong Yi Yao Za Zhi,*" "*Chinese Journal of Integrative Medicine*". We informally express each row as an expression $C \oplus Q = S(s_1, s_2, \ldots, s_p)$ to describe the problem more easily. The same disease or the same syndrome usually occurs in more than one clinical cases. The following expressions demonstrate the possible scenarios referring to the clinical cases:

$$\begin{aligned} C_1 \oplus Q &= S_1\left(s_1, s_2, \ldots, s_p\right), \\ C_2 \oplus Q &= S_2\left(s_1', s_2', \ldots, s_q'\right), \end{aligned} \tag{1}$$

$$\begin{aligned} C \oplus Q_1 &= S_1\left(s_1, s_2, \ldots, s_p\right), \\ C \oplus Q_2 &= S_2\left(s_1', s_2', \ldots, s_q'\right). \end{aligned} \tag{2}$$

In TCM diagnosis, though it is difficult to build a complete model to understand the complicated theory, but it is acknowledged that the emergence of pathological patterns (such as syndromes, disease, and symptoms) shows certain laws. The frequency of the appearance of the same symptom shows the strength of connection to syndromes or diseases. The more the symptom appears in the disease or syndrome related expressions, the firmer the connection is. Thus, we transform the frequency calculation problem of symptoms to a set intersection problem.

The intersection problem in processing the clinical data can be defined as finding the relatively small but compact set of symptoms for each disease and syndrome. For the expressions (1), we should conclude that the syndrome Q is coherently related to the mixed set of symptoms $S = S_1 \bigcap S_2$; similarly, the disease C can be mapped to the set of symptoms $S = S_3 \bigcap S_4$ concluded from the expressions (2). It is obvious that the elements found are the intersection of both sets. These expressions are collected and formalized directly from each row of clinical database, and the most relevant symptoms are concluded with the intersected sets.

Consider the problem of computing the intersection of k sets of various size, and the value of k might varies a lot for different test cases (different syndromes or diseases). In 2000, Demaine et al. [18] introduced a new intersection algorithm, named *Adaptive*, which intersects all the sets in parallel so as to compute the intersection in time proportional to the shortest proof of the result set. They studied that in the context of Internet information queries and text database systems. If queries are composed of words, and for each keyword a sorted set of references to entries in the database is precomputed, then the set of data entries in the database matching all the keywords of a query is the intersection of the sorted sets corresponding to each keyword. After that, Jérémy have done an experimental investigation [19] of set intersection algorithms and extend the optimal algorithm to the *t-threshold problem* [20], which consists in finding the elements which consists in finding the elements which are in at least t of the k sets.

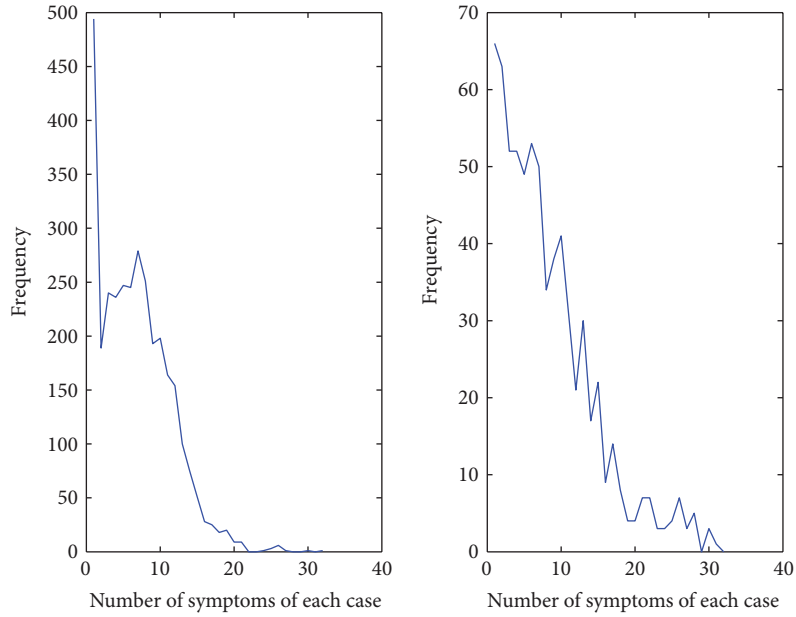
Although in general case, sets $S = \{S_1, S_2, \ldots, S_n\}$ are arbitrary, an important case is when each set S_n is already

Input: multi-set $A = \{S_1, S_2, \ldots, S_k\}$
Output: T
if $A.size() == 0$ **then**
  return null;
end
else if $A.size() == 1$ **then**
  return A. get (0);
end
t = generateRatioNum (A. size());
Let $M \leftarrow t$th largest elements of the multi-set $\{A_1[n_1], A_2[n_2], \ldots, A_k[n_k]\}$;
Let $H = \{A_1[1], A_2[1], \ldots, A_k[1]\}$;
Let $m \leftarrow t$th element of H;
Mark green all the sets A_i such that $A_i[1] = m$, and remove all copies of m from H;
Mark red all the sets A_i such that $A_i[1] > m$;
Mark white all the sets A_i such that $A_i[1] < m$, and remove $A_i[1]$ from H;
While $m < M$ **do**
  if *t sets are green or $k - t + 1$ sets are red* **then**
    if *t sets are green* **then**
      $T_t \leftarrow T_t \bigcup m$
    end
    Take $t - 1 -$ # (white sets) of the green sets and mark them white;
    For each remaining green set A_i, insert in H the first element of A_i which is $> m$, and change the mark of A_i to red;
    Let $m \leftarrow \min H$;
    change the mark to green for all the sets which have m as a representative in H, and remove m from H;
  end
  Let A_i be the next white set;
  if $m \in A_i$ **then**
    Mark A_i in green;
  end
  else
    Insert in H the first element of A_i which is strictly larger than m, and mark A_i in red;
  end
end

Algorithm 1: Minimum symptom set extraction.

in order. In this case, the performance of set intersection algorithm can be enhanced. However, this is not optimal for all possible cases. In fact, if m is small, it is better to do m binary searches obtaining an $O(m \log n)$ algorithm [21]. When m and n are both large, it is necessary to look up most of the elements as effective as possible. According to [20], any algorithm for the intersection problem must certify that the output is correct: first, it must certify that all the elements of the output are indeed elements of all the sets; second, it must certify that no element of the intersection has been omitted by exhibiting some inequalities which imply that there can be on other element in the intersection. Therefore, they prove that any deterministic algorithm for intersection problem must take at least $\Omega(\delta \sum_i \log(n_i/\delta))$. Here, δ is defined as the difficulty of an instance. Given k ordered sets $A_1, A_2, \ldots, A_k$ of sizes $n_1, n_2, \ldots, n_k$, the variant t-threshold problem consists in computing the set of elements which are present in at least t of k sets. In particular, t varies from the magnitude of k, namely, the cardinality of the input sets. We applied an altered adaptive intersection algorithm (Algorithm 1) based on t-threshold [20] to construct a feature selection library. We simplify the result of small data set and eliminate the doubling search process for unnecessary redundance. The content of the generated diagnostic library is a collection of expressions in the form of $D \approx S$, in that D represents a kind of disease or syndrome, and S represents a set of most relevant symptoms. Usually the clinical data is incomplete and noisy, which may causes inadequate mining results. For example, the number of symptoms in each case varies over a wide range, from to 1 to 32, as shown in Figure 2, and the average quantity of symptoms for each row in the sample database is 6.787. After the process of data preprocessing, we generate 705 $D \approx S$ mapping relations with an average length of 8.21 of symptoms for each case. It is obvious that the length of symptoms becomes more well distributed, with less frequency near 0, more frequency near the center of the range.

Through the set intersection calculation of original clinical records, each syndrome or disease is associated with a relatively minimum set of symptoms. The reason why both syndromes and diseases matter to the TCM diagnosis is that they will both be considered in the process of clinical diagnosis, and treatments will be taken under the classification results.



(a) Original sample database (b) Diagnostic library

Figure 2: Number of clinical cases by length.

6. United System of Syndrome Differentiation

6.1. Ontology-Based Relation Refinement. To overcome the heterogeneity of information sources of different domains and attribute volumes, ontologies based on conceptual graphs are integrated to represent component information, such as the field labels that contain the same kind of information from different databases [22]. The use of ontologies for the explication of implicit and hidden knowledge is a possible approach to overcome the problem of semantic heterogeneity, which eliminates the inconsistency of information resources and distributed information bases.

Semantic heterogeneity considers the content of an information item and its intended meaning. According to [23], there are three main causes for semantic heterogeneity: *confounding conflicts*, *scaling conflicts*, and *naming conflicts*. Naming conflicts occur when naming schemes of information differ significantly, such as homonyms and synonyms.

With respect to the integration of data sources, ontologies can be used for the identification and association of semantically corresponding information concepts. We built a single ontology as one global ontology providing a shared vocabulary for the specification of the semantics. All information sources are related to one global ontology. As mentioned in Section 2, we defined an *equivalent class* with no restriction as a synonym for syndromes, a *subClassOf* subsumption relation as a categorical definition, and a set of *restrictions* with object property relation as a customized relation "subsume." The capability of using ontology will be enhanced by perform inference on it and generating new facts.

The integration process between ontology and database is demonstrated in Figure 3. In this figure, the hierarchy of syndromes consists the most important theoretical base for medical diagnosis. Relations between syndromes defined in ontology are integrated with rows of the clinical database in a directed way. We use strict text matching strategy to map the ontological definitions to database columns, specifically the "`syndromes`" column. We propose four strategies or steps for data integration after the first step of data preprocessing in Section 5; here S stands for the input set of symptoms, and T stands for currently matching symptom set.

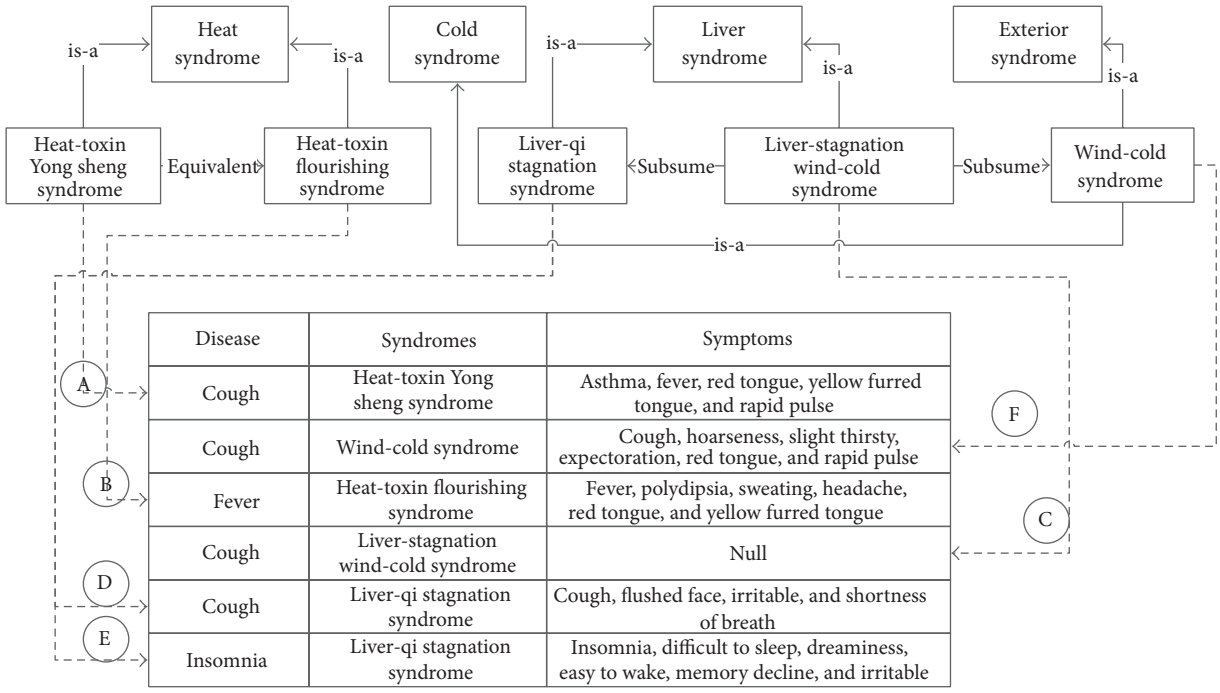
(1) $S \supset T$: the syndrome set T only matches part of the symptoms S, it means that the input syndromes may match some other syndrome with higher similarity or match the syndrome which *"subsume"* the syndrome T (*A tcm:subsume syndrome(T)*).

(2) $S \subset T$: the symptom set T is much larger that the input set S, which means a smaller set may be more perfect, then we should consider the super classes of *syndrome(T)* (*syndrome(T)* $\supset$ *superclass(syndrome(T))*).

(3) $S = T$: the symptom set S matches perfectly with S, *syndrome(T)* or its equivalent syndrome should be a perfect answer.

(4) $S \cap T$: the input symptom set S has intersection with the possible symptom set T with a degree of matching above certain threshold, the unmatched part of input sets are recalculated for most relevant local result. The final result is consisted of the local results as combined syndromes. It is quite common in TCM that syndromes can happen together with one disease.

Here, *syndrome(T)* stands for the specific syndrome relates to the symptom set T. Instead of iteratively calculating the corresponding symptoms for each syndrome, we would rather



Disease	Syndromes	Symptoms
Cough	Heat-toxin Yong sheng syndrome	Asthma, fever, red tongue, yellow furred tongue, and rapid pulse
Cough	Wind-cold syndrome	Cough, hoarseness, slight thirsty, expectoration, red tongue, and rapid pulse
Fever	Heat-toxin flourishing syndrome	Fever, polydipsia, sweating, headache, red tongue, and yellow furred tongue
Cough	Liver-stagnation wind-cold syndrome	Null
Cough	Liver-qi stagnation syndrome	Cough, flushed face, irritable, and shortness of breath
Insomnia	Liver-qi stagnation syndrome	Insomnia, difficult to sleep, dreaminess, easy to wake, memory decline, and irritable

Figure 3: The integration of TCM domain knowledge and the clinical cases in the database.

postpone the integration of metaknowledge and databases, until the actual implementation of query answering.

6.2. Fuzzy Diagnosis. When a patient comes, not knowing what kind of disease he or she has, must summarize his or her symptoms, including external appearance such as red face, yellow tongue fur, and detected symptoms including string pulse. When these symptoms are submitted to the ODS, they will be compared to the symptom-set list with similarity match algorithm to find the most similar set of disease A and syndrome B. After this, the ontology will be involved to make relation refinement according to the hierarchical information such as equivalence, subsumption, inheritance. The symptom set of the closely related syndromes of B will be recalculated. At last, the system will give the most relevant syndrome(s) and disease. In general, the main goal is to identify the most relevant diagnostic result (including disease and syndromes) for an input of symptoms. Due to the randomness of user-input symptoms, the input should be carefully matched with technical terms in the processed clinical database by measuring similarity.

Basically, since we are not ready for semantic text match, the measure of symptom set similarity of the system is based upon simple text match by identifying the long common subsequence [24]. The longest common subsequence measures the similarity letter by letter, which ranges from 0 to 1. The information of *weight* tells the ratio of matched string sequence over the longer one of the two pieces of texts. Frequency of full-text match (when *weight* equals to 1) is counted, and we multiply the sum of *weight* for each word by the frequency, then dividing the size of the symptom list (number of symptoms in the list). The similarity *sim* between input symptoms and each specific group of symptoms in the table P is a float number obtained during each comparison. The input set of symptoms is matched with each clinical case for each kind of diseases and syndromes. Both the disease and the syndrome of the final answer will be separately found out once the cases with the largest similarity pop out one has.

$$\begin{aligned} \text{sim}\left[\text{input}, \text{list}\right] &= \frac{\text{num}\left(\text{weight} = 1\right) \times \sum \text{weight}}{\text{list} \cdot \text{size}\left(\right)}, \\ \text{weight} &= \frac{\text{length}\left(\text{lcs}\right)}{\text{length}\left(\max\left(a, b\right)\right)}. \end{aligned} \tag{3}$$

The process of obtaining diagnostic results is actually a process of sorting and filtering. An array of input symptoms I will be compared to each record of list L in the table P based on (3), in order to find an initial diagnostic result. The *fuzzy* here means that we assign a weighted value for both word matching and set matching. *Fuzzy word matching* inherited from basic text matching method defines a truth value that ranges in degree between 0 and 1, which measures the similarity between input words (user input) and technical terms (in the database). *Fuzzy set matching* takes both the sum of weights and the number of symptoms into account. We do not consider the fuzziness of descriptions in symptoms, such as a little pain, since it is another research problem referring to fuzzy logic. Several researches have been done to apply the fuzzy logic proposed by Zadeh [25] in 1965 into medical diagnosis [26].

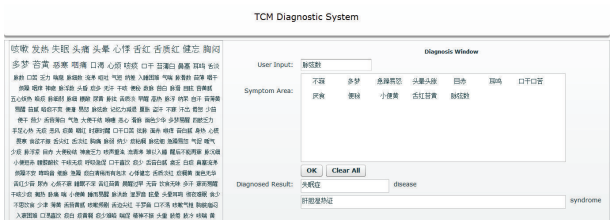


Figure 4: The snapshot of Ontology-oriented Diagnostic System (ODS).

7. Results and Validation

Based on the methods we proposed, we implemented an interactive medical diagnostic platform deployed on web browser. Online diagnostic system can provide users with real-time clinical suggestions rather than actual medical diagnosis due to the lack of knowledge and data within computer-aided system. For the user interface design, we apply Adobe Flex technology to build a flexible web application, which enables rich user interaction and query answering. *Flex* is a software development kit (SDK) released by Adobe Systems for the development and deployment of cross-platform Rich Internet Applications (RIA) based on the Adobe Flash platform.

In detail, our user interface is consisted of three parts: *tag cloud panel*, *input panel*, and *output panel*. Initially the user is presented with the most frequentlyoccurred symptoms in the left tag cloud panel, which provides users with an overview of all frequentlyoccurred symptoms and enable easy selection just by clicking on the word itself. The tag cloud panel guides the user in the construction of a tag cloud by means of frequency ranking. Largest texts means bigger chances to appear, smaller vice versa. The selected symptoms would be added into the input panel. Alternatively, the user can type vulgar symptoms into the input area. Entered symptoms would also be added into the input panel. Finally, all the chosen symptoms will be listed out in the right-top symptom area. The patient can fetch the diagnostic result whether he plans to do some tests on the system or have an actual requirement for medical help. The calculation for the diagnostic result is strictly behaved through our proposed methods. After the work flow, the expected disease and syndrome are listed in the output panel. A snapshot of the user interface is demonstrated as in Figure 4.

The user interaction was designed using a prototype-based interactive process, which aims to find problems and clarify the design principles as early as possible. It turns out that users prefer simplicity and vividness of our prototype, but for a diagnostic system, the accuracy of diagnostics is most important. Three metrics are imported to measure the accuracy, feasibility, and performance of the system.

7.1. Accuracy. Due to the diversity of medical domains, there are no open benchmark for testing the accuracy of diagnostic results, especially for TCM. Thus we design a metric method for accuracy: A–E value metric (A for "Very precise," B for "precise," C for "tolerable," D for "erroneous," and E for "wrong").

Table 1: Main characteristics of *SynOnt* ontology.

Number of classes (syndromes)	391
Number of subclass axioms	483
Number of equivalent class axioms	254
Number of subsumed class axioms	122

Table 2: Experimental Results of ODS diagnostics (total accuracy/syndrome accuracy).

Number	Nonontology diagnosis	Ontology-oriented diagnosis
1	A/A	A/B
2	C/B	A/A
3	C/C	A/B
4	C/D	A/A
5	B/A	B/B
6	C/D	B/B
7	E/E	A/A
8	A/A	A/A
9	B/A	B/C
10	C/B	A/A
11	C/B	A/A
12	B/A	C/B

The system was evaluated by two medical experts from CATCM, who provided 12 test cases and scored the diagnostic result in A–E. All the test cases are directly selected from the authoritative textbook of Chinese Internal Medicine, which has no direct connection with the sample database. The scoring results are listed as in Table 1, in which each domain expert gives both total accuracy (disease plus syndrome diagnosis) and syndrome accuracy, because the disease diagnosis concludes on the patient's health condition, and syndrome diagnosis matters to the accuracy of syndrome differentiation. The system can achieve good diagnostic decisions, in which the overall diagnostic accuracy (O) scores distributes as 67%A, 25%B, 8%C, and no votes for D and E. The relative score (R) measures the accuracy of single-syndrome matching, which distributes over 50%A, 42%B, 8%C, and no votes for D and E, as shown in Table 2.

We found that most users accepted our system with positive reactions and considered the system generally useful to help medical practitioners with clinical decision making. Furthermore, we compared the diagnostic results using only minimum set extraction method and the complete ontology-oriented method, and it turned out that 6 over 12 test cases got much better results using the ontology oriented method.

7.2. Performance. The performance of our system is also measured, which estimates the efficiency of our algorithms: (1) the symptom calculation of diagnostic library, (2) integration algorithm of heterogeneous data sources.

It is known that Algorithm 1 computes the t-threshold calculation of multisets of difficulty δ in $O(t\delta \log k \log n)$. From Figure 5, we can see that the computation cost of each

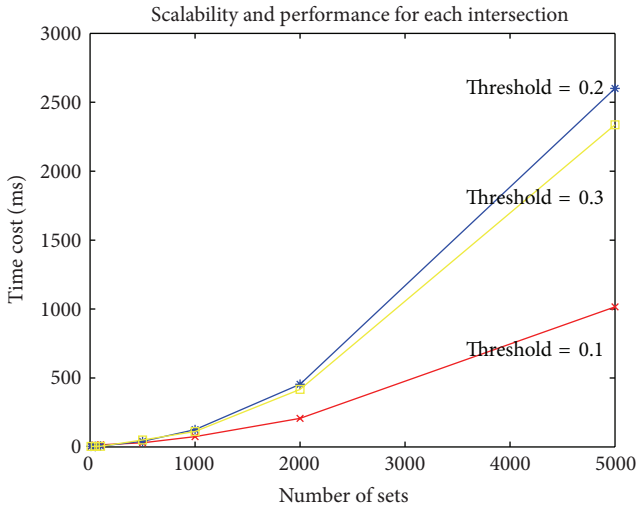


Figure 5: The performance curve for each intersection.

set intersection calculation is raising by the number of sets involved, not to mention doing t-threshold multiple times. However, when the threshold increases, which means the condition turns to be stricter, the time cost will fall. The integration strategy checks the mathematic relation between inputs and primary result and matches extended choices by calculating similarity with low cost.

7.3. Scalability. Since the practical basis of our system is mainly based on the intersection calculation of clinical cases, the generated diagnostic library could vary due to different clinical cases in terms of content and volume. Thus, the computation cost of the system is measured with the increasing number of clinical cases. The number of clinical cases will only affect the scalability of set intersection algorithm.

The computation cost raises fast with the increasing number of clinical cases, such that it is bearable only if it is offline preprocessing when a large number of cases and disorder concepts involve in.

8. Conclusion and Discussion

In this paper, we implemented an integrated diagnostic system for Traditional Chinese Medicine, which learns diagnostic principles from prior clinical experience to provide clinical decision-making services. As an alternative source of health care, TCM is interpreted as to have intangible connections between human and nature rather than anatomical parts, which leads to complex semantic inclusion relations. We have done some researches on expressing biological facts into ontological statements, by constructing domain ontology as OWL and RDF models. Semantic technologies show great potential on querying and reasoning the knowledge of TCM [27]. The ontology assisted diagnostic system interprets the correspondence between symptoms and syndromes in an integrated method of minimum set mapping and ontology refinement, instead of static rules which are difficult to conclude. Web users could access the online user interface and fetch a diagnostic result according to the specific input symptoms.

Over 10,000 patient records are collected from curated data sources to constitute a sample database, in which attributes could be grouped into three columns: disease, syndrome, and symptoms. We run a minimum symptom set extraction algorithm for each syndrome and disease so as to generate a small featured set of symptoms for each of them to form a statistical mapping table. Since there are naming conflicts and inclusion association in TCM terms, we codify the hierarchy of syndromes and logical relations (such as class inheritance, equivalence, and inclusion) extracted from literature with the help of TCM experts into a syndrome ontology model. The finalized diagnostic result is obtained after a relation refinement process over the mapping table based on the ontology information. We process text similarity matching and ontology refinement for the input data to find the most relevant result. To verify the accuracy of our system, we invite medical experts in CATCM to provide 12 test cases and score the result of each case using a grading metric.

There are a lot to be further investigated. The key consideration of building an intelligent diagnostic agent is to understand the essence of TCM medical theories, which may consists of philosophical ideas about five phases and body, treatment principles, diagnostic methods, and so forth. The most import component of our diagnostic system is the mapping table (diagnostic library), which learns diagnostic principles by t-threshold intersection algorithm. The performance of the algorithm raises fast with the increasing amount of clinical cases. The performance problem may be resolved by distributing sample data into several commodity machines to reduce computation cost, probably by increment strategies.

Acknowledgment

The work of the authors is funded by NSFC61070156, national 863 program China Cloud Initiative, and the Fundamental Research Funds for the Central Universities.

References

[1] H. Wang, "A computerized diagnostic model based on naive bayesian classifier in traditional chinese medicine," in *Proceedings of the 1st International Conference on BioMedical Engineering and Informatics (BMEI '08)*, vol. 1, pp. 474–477, IEEE, May 2008.

[2] X. Wang, H. Qu, P. Liu, and Y. Cheng, "A self-learning expert system for diagnosis in traditional Chinese medicine," *Expert Systems with Applications*, vol. 26, no. 4, pp. 557–566, 2004.

[3] S. Mukherjea, "Information retrieval and knowledge discovery utilising a biomedical Semantic Web," *Briefings in Bioinformatics*, vol. 6, no. 3, pp. 252–262, 2005.

[4] B. L. Humphreys and D. A. B. Lindberg, The unified medical language system.

[5] K. Spackman, K. Campbell, R. A. Côté et al., "SNOMED RT: a reference terminology for health care," in *Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association*, p. 640, 1997.

[6] A. Rodríguez-González, G. Hernández-Chan, R. Colomo-Palacios et al., "Towards an ontology to support Semantics enabled diagnostic decision support systems," *Current Bioinformatics*, vol. 7, no. 3, pp. 234–245, 2012.

[7] Á. García-Crespo, A. Rodríguez, M. Mencke, J. M. Gómez-Berbís, and R. Colomo-Palacios, "ODDIN: ontology-driven differential diagnosis based on logical inference and probabilistic refinements," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2621–2628, 2010.

[8] M.-J. Huang and M.-Y. Chen, "Integrated design of the intelligent web-based Chinese Medical Diagnostic System (CMDS)—systematic development for digestive health," *Expert Systems with Applications*, vol. 32, no. 2, pp. 658–673, 2007.

[9] Y. Mao and A. Yin, "Ontology modeling and development for Traditional Chinese Medicine," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, pp. 1–5, IEEE, October 2009.

[10] S. Lukman, Y. He, and S. C. Hui, "Computational methods for traditional Chinese medicine: a survey," *Computer Methods and Programs in Biomedicine*, vol. 88, pp. 283–294, 2007.

[11] D. L. Rubin, N. H. Shah, and N. F. Noy, "Biomedical ontologies: a functional perspective," *Briefings in Bioinformatics*, vol. 9, no. 1, pp. 75–90, 2008.

[12] F. Manola and E. Miller, RDF Primer: W3c recommendation, Decision Support Systems.

[13] D. L. Mcguinness and F. V. Harmelen, OWL Web Ontology Language Overview, World Wide Web.

[14] E. Antezana, M. Kuiper, and V. Mironov, "Biological knowledge management: the emerging role of the Semantic Web technologies," *Briefings in Bioinformatics*, vol. 10, no. 4, pp. 392–407, 2009.

[15] G. Sun, The Basic Theory of Chinese Medicine, China Press of Traditional Chinese Medicine, 2005.

[16] G. Zhu, Diagnostics of Chinese Medicine, China Press of Traditional Chinese Medicine, 2007.

[17] Z. Zhou, Chinese Internal Medicine, China Press of Traditional Chinese Medicine, 2007.

[18] E. D. Demaine, A. López-Ortiz, and J. I. Munro, "Adaptive set intersections, unions, and differences," in *Symposium on Discrete Algorithms*, pp. 743–752, 2000.

[19] J. Barbay, A. López-Ortiz, T. Lu, and A. Salinger, "An experimental investigation of set intersection algorithms for text searching," *ACM Journal of Experimental Algorithmics*, vol. 14, 2009.

[20] J. Barbay and C. Kenyon, "Adaptive intersection and t-threshold problems," in *Symposium on Discrete Algorithms*, pp. 390–399, 2002.

[21] R. A. Baeza-yates, "A fast set intersection algorithm for sorted sequences," in *Combinatorial Pattern Matching*, pp. 400–408, 2004.

[22] H. Wache, T. Vögele, U. Visser et al., "Ontology-based integration of information-a survey of existing approaches," in *Proceedings of the IJCAI Workshop on Ontologies and Information Sharing*, 2001.

[23] C. H. Goh, Representing and reasoning about semantic conflicts in heterogeneous information systems, 1997.

[24] D. S. Hirschberg, "Algorithms for the longest common subsequence problem," *Journal of the Association for Computing Machinery*, vol. 24, no. 4, pp. 664–675, 1977.

[25] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, pp. 338–353, 1965.

[26] S. K. De, R. Biswas, and A. R. Roy, "An application of intuitionistic fuzzy sets in medical diagnosis," *Fuzzy Sets and Systems*, vol. 117, no. 2, pp. 209–213, 2001.

[27] H. Chen, Y. Wang, H. Wang et al., "Towards a semantic web of relational databases: a practical semantic toolkit and an in-use case from traditional Chinese medicine," in *Proceeding of the 5th international conference on The Semantic Web (ISWC '06)*, vol. 4273 of *Lecture Notes in Computer Science*, pp. 750–763, 2006.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 650671, 13 pages
http://dx.doi.org/10.1155/2013/650671

Research Article

The iOSC3 System: Using Ontologies and SWRL Rules for Intelligent Supervision and Care of Patients with Acute Cardiac Disorders

Marcos Martínez-Romero,[1] José M. Vázquez-Naya,[2] Javier Pereira,[1] Miguel Pereira,[3] Alejandro Pazos,[2] and Gerardo Baños[3]

[1] *IMEDIR Center, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain*
[2] *Department of Information and Communication Technologies, Computer Science Faculty, University of A Coruña, 15071 A Coruña, Spain*
[3] *Service of Anesthesiology, Resuscitation and Intensive Care, Meixoeiro Hospital, Meixoeiro s/n, 32600 Vigo, Spain*

Correspondence should be addressed to Marcos Martínez-Romero; marcosmartinez@udc.es

Received 26 October 2012; Revised 25 December 2012; Accepted 3 January 2013

Academic Editor: Alejandro Rodríguez González

Physicians in the Intensive Care Unit (ICU) are specially trained to deal constantly with very large and complex quantities of clinical data and make quick decisions as they face complications. However, the amount of information generated and the way the data are presented may overload the cognitive skills of even experienced professionals and lead to inaccurate or erroneous actions that put patients' lives at risk. In this paper, we present the design, development, and validation of iOSC3, an ontology-based system for intelligent supervision and treatment of critical patients with acute cardiac disorders. The system analyzes the patient's condition and provides a recommendation about the treatment that should be administered to achieve the fastest possible recovery. If the recommendation is accepted by the doctor, the system automatically modifies the quantity of drugs that are being delivered to the patient. The knowledge base is constituted by an OWL ontology and a set of SWRL rules that represent the expert's knowledge. iOSC3 has been developed in collaboration with experts from the Cardiac Intensive Care Unit (CICU) of the Meixoeiro Hospital, one of the most significant hospitals in the northwest region of Spain.

1. Introduction

An Intensive Care Unit (ICU), sometimes called Critical Care Unit, is a special area in the hospital used to treat the most critically ill patients. ICUs are amongst the busiest and most high-pressure units in a hospital, because ICU patients require constant supervision and treatment, either because they are recovering from a major operation, have an acute illness, or have been injured in a severe accident. Medical staff in ICUs are specially trained in procedures like acute pain management and how to use specialist equipment. They also need to be able to make quick decisions in high-stress situations and stay calm, when lives are at stake.

Life of patients in the ICU depends largely on the wisdom of these decisions. Nevertheless, experts in health environments are frequently exposed to long working hours, extended days, and shift-work schedules besides high workload and psychological strain [1]. As a consequence, the human factor sometimes leads to imprecisions or mistakes in the decision making process that represent a barrier to the optimal recovery of patients. It has been reported that critically ill patients admitted to an ICU experience, on average, 1.7 medical errors each day, and numerous patients suffer a potentially life-threatening error during their stay in the ICU [2, 3]. The most frequent errors are due to the incorrect administration of medications [4, 5]. We believe that the application of Artificial Intelligence techniques to provide decision support at ICUs can help doctors to monitor and care patients, decreasing the number of treatment mistakes and improving the recovery process.

This work presents iOSC3 (intelligent Ontology-based System for Cardiac Critical Care). iOSC3 is a decision support system designed to supervise and treat patients affected by acute cardiac disorders. In hospitals with limited resources, this type of patients are treated at the general ICU. Nevertheless, when possible, they are treated at Cardiac Intensive Care Units (CICUs), which are a special type of ICU exclusively addressed to recover patients with cardiac illnesses. The system has been designed in collaboration with the CICU of the Meixoeiro Hospital (http://chuvi.sergas.es), one of the most significant hospitals in the northwest region of Spain. On the basis of the patient's vital signs at a particular time and an ontology that formalizes the expert's knowledge, the system provides a recommendation about the treatment that should be administered to achieve the fastest possible recovery.

1.1. Cardiac Intensive Care Units. The Cardiac Intensive Care Unit (CICU) is an area of the hospital reserved for treating patients with life-threatening cardiac medical conditions. This can include recovery from open heart surgery, heart attack recovery, or different medical issues that put the heart at risk. Medical staff in CICUs have training in specialty areas, including critical care, cardiology, and cardiac anesthesia. Their knowledge, jointly with sophisticated monitoring and support equipment, enables them to recognize problems quickly and respond effectively.

Each patient in the CICU is connected to a TV-like screen (patient monitor) which continuously acquires and displays different measures that represent body activities, such as temperature, heart rate, mean arterial pressure, and cardiac index. The patient is also connected to one or several infusion pumps, containing the drugs used to stabilize him/her. The drugs used at the CICU must be supplied judiciously and with a goal-directed approach. They are classified into two groups: vasodilators and amines (vasopressors). On the one hand, vasodilators, as the name implies, relax the smooth muscle in blood vessels, which causes the vessels to dilate and decreases blood pressure. Examples of vasodilator drugs commonly used at the CICU are nitroglycerin and nitroprusside. On the other hand, amines as dobutamine, noradrenaline, adrenaline, and dopamine have vasoconstriction properties. They increase blood pressure, which increases organ perfusion pressure and preserves distribution of cardiac output to the organs. They also improve cardiac output and oxygen delivery by decreasing the compliance of the venous compartment and thus augmenting venous return.

According to the patient's condition, CICU physicians adjust the infusion rate of vasodilators and amines in order to stabilize him/her. Sometimes, they have to integrate several rapidly changing physiologic parameters into a clear and qualitative mental image of a patient's current state and take a decision about the amount of the drugs to be administered in a short period of time. As previously explained, the system proposed in this paper is aimed to provide support to physicians in these complex decisions. The system has been developed in collaboration with experts from the CICU of the Meixoeiro Hospital (Vigo, Spain), and it has been tested in the CICU of such hospital, which is a 10-bed unit designed, equipped, and staffed to provide interdisciplinary critical care to those patients with cardiac disorders (e.g., myocardial infarction, acute coronary syndrome, congestive heart failure, pre- and postcardiac transplants, high-risk cardiac arrhythmias, etc.).

1.2. Ontologies in Biomedical Research. Towards the end of the 20th and beginning of the 21st centuries, and especially since the Semantic Web was conceived [6], the term "ontology" (or ontologies) gained usage in Computer Science to refer to a research area in the subfield of Artificial Intelligence mainly concerned with the semantics of concepts and with expressive processes in computer-based communications. In Computer Science, ontologies are a technique used to represent and share knowledge about a domain by modeling the things in that domain and the relationships between those things [7]. Ontologies are represented using standard, machine-processable languages (e.g., RDF [8] and OWL [9]), and they are mainly used for communication between people and organizations by providing a unified terminology that allows to reach a common level of understanding or comprehension within a particular domain.

In the biomedical field, ontologies have increasingly become an established method to represent and communicate the huge amount of knowledge about genes, diseases, biomedical processes, and so forth that has been generated during the last years [10]. Biomedical ontologies are considered crucial pieces in the development of informatics applications in several areas, such as knowledge-based decision support, terminology management, and systems interoperability and integration [11]. As a consequence, multiple biomedical ontologies have been developed and maintained, which are stored in large-scale ontology repositories available for researchers. The most popular repository of biomedical ontologies is the NCBO's BioPortal [12], a web-based, open resource that contains more than 300 ontologies with knowledge related to different biomedical topics (anatomy, gene products, immunology, phenotype, etc.) in different organisms (human, plant, mouse, microbe, etc.).

In the medical domain, ontologies are key to reuse the large amount of complex information that is involved in many health care activities. They are used to build systems for purposes such as data annotation, information retrieval, and natural-language processing, but they are particularly useful to build knowledge-based systems that provide decision support in health care. This type of systems are generally dependent on large volumes of domain knowledge, which is extremely expensive and difficult to capture and formalize [13]. By means of ontologies, this knowledge can be represented in an application independent manner; so, it can be reused in new systems without additional knowledge extraction and development effort.

In this work, we have constructed an ontology (called Critical Cardiac Care Ontology, or C3O) that contains the knowledge used by CICU physicians to diagnose and treat patients. This ontology, which is presented in Section 3, is used by the iOSC3 system to analyze the patient's condition

and to provide decision support about the treatment that should be administered to recover him/her.

1.3. Related Work. First contributions in the field of expert systems were made by the Artificial Intelligence community in the late 50s and early 60s, when several programs aimed at general problem solving were written. However, the first generation of clinical decision support approaches date back to the early 1970s. Popular examples are AAPHelp [14], designed to support the diagnosis of acute abdominal pain, INTERNIST-I [15], a rule-based expert system aimed to provide diagnostic support in the domain of internal medicine, and MYCIN [16], a very powerful system for diagnosing blood infections and recommending their antibiotic therapies, which has been described as the first convincing demonstration of the power of the rule-based approach in the development of robust clinical decision support systems [17]. Since the 70s, several expert systems have been proposed to solve diverse problems in medical domains, including intensive care environments. Some recent examples are an expert system for electroencephalogram monitoring in the pediatric ICU [18], an ontology-driven medical diagnosis system [19], a fuzzy logic system to regulate mean arterial pressure [20], an expert system for detection of breast cancer [1], a rule-based solution that applies semantic web techniques to ensure patient safety during breast cancer surgery [21], a system for improving specificity of alarms in critical care environments [22], a hybrid approach using case-based reasoning and rule-based reasoning for decision support in ICUs [17], and a recommendation system for antidiabetic drugs selection [23].

Table 1 summarizes the main features of iOSC3 and recent related experts systems. The systems have been classified according to the categories proposed in [24]. Despite previous work, to the best of our knowledge, no expert system has been developed to monitor and control patients at CICUs. In addition, Table 1 shows that many of the systems developed so far use traditional knowledge representation techniques that are not adequate for sharing the expert's knowledge with other professionals and to reuse it in other similar systems.

With respect to existing ontologies for the intensive care domain, during the last decade, some researchers have worked to structure and standardize existing knowledge by means of ontologies (e.g., [25, 26]). However, despite previous efforts, there is no ontology addressed to cover the cardiac intensive care domain. With this in mind, the expert system we present in this paper and the ontology that provides the underlying expert's knowledge constitute an innovate contribution in the field of medical expert systems.

2. The iOSC3 System

The system has been developed in Java technology by means of the Eclipse IDE and following the Unified Software Development Process [27]. The system architecture and workflow will be described in the sections which follow.

2.1. System Architecture. The overall architecture of iOSC3 is shown in Figure 1. Each constituent will now be described in further detail.

2.1.1. CICU Devices. Patients in CICUs are connected to a monitor, which is a TV-like screen that continuously acquires and displays different measures that represent body activities, such as heart rate, mean arterial pressure, and cardiac index. They are also connected to one or several infusion pumps, depending on their medical condition. These pumps are specialty devices designed to deliver controlled doses of medications that are used when it is otherwise impossible to treat a patient at the prescribed times and quantities, such as in minute quantities smaller than what a drip system can deliver. In the CICU of the Meixoeiro Hospital, the specific models of devices that are being used are Philips IntelliVue MP Series monitors and Alaris TIVA infusion pumps.

2.1.2. Communication APIs. In order to enable the system's interaction with the CICU devices, it has been necessary to develop two communication APIs: the Monitor Communication API, which allows the system to establish a communication with Philips IntelliVue MP Series monitors to obtain the values of the patient's vital signs, and the Pumps Communication API, which makes it possible to obtain the state of the Alaris IVAC infusion pumps (current infusion rate and infusion state) and also to change it.

Philips IntelliVue patient monitors use a well-established communication protocol called Data Export Protocol, which is a connection-oriented, message-based request/response protocol that works on top of the standard UDP/IP transport protocol [28]. By means of the Data Export Protocol, data from the monitor can be transferred via the Local Area Network (LAN) Interface or Medical Information Bus (MIB/RS232) Interface to an external computer.

We have developed an API (Monitor Communication API) that implements the Data Export Protocol in order to obtain data from the IntelliVue monitor via the LAN interface. The iOSC3 system is connected to the IntelliVue monitor using a standard unshielded LAN cable with an RJ45 connector. The network IP address is automatically configured with the standard Bootstrap Protocol (BootP).

The communication with the Alaris syringe pumps is achieved by means of the Alaris Pump Communications Protocol [29]. This protocol is used for the pump models GS, GH, CC, and TIVA with software version V1.5.10 (or V1.6.2 for TIVA models) and above. The communications model is a point-to-point connection between two communicating parties (a client and a server), and it has been designed to support two types of connection: conventional RS232 serial interface and IrDA infrared specification.

Our Pumps Communication API allows to communicate the system with the Alaris pumps by means of the RS232 serial interface. It enables the system to obtain the state of the infusion pump, as well as to modify it.

2.1.3. Expert System. The decision making process provided by iOSC3 is supported by a specialized expert system that is

Table 1: Comparison of iOSC3 with related medical expert systems. Each system is classified according to the categories proposed in [24]. The last column summarizes the main techniques or technologies used for knowledge representation and processing.

System	Category	Application	Techniques/technologies
iOSC3	Ontology-based, rule-based	Decision support in cardiac ICUs	OWL ontology, SWRL rules, Pellet reasoner
Chen et al. (2012) [23]	Ontology-based, rule-based	Antidiabetic drugs selection	OWL ontology, SWRL rules, JESS engine
ODDIN (2010) [19]	Ontology-based, rule-based, probabilistic	Differential diagnosis in medicine	OWL ontology, Jena rules
Nocedal et al. (2010) [21]	Rule-based	Breast cancer treatment	OWL ontology, inference rules
Kumar et al. (2009) [17]	Case-based, rule-based	Clinical decision support in ICUs	Rules in XML format
Blum et al. (2009) [22]	Intelligent agents, rule-based	Improving physiologic alarms in critical care	Inference engine implemented using a stored SQL procedure
Karabatak and Ince (2009) [1]	Association rules, neural network	Breast cancer detection	Association rules for feature extraction and multilayer perceptron for intelligent classification
Si et al. (1998) [18]	Fuzzy logic, neural network	Electroencephalogram monitoring in pediatric ICUs	Statistical comparison of features, fuzzy logic for feature classification, and neural networks for EEG assessment

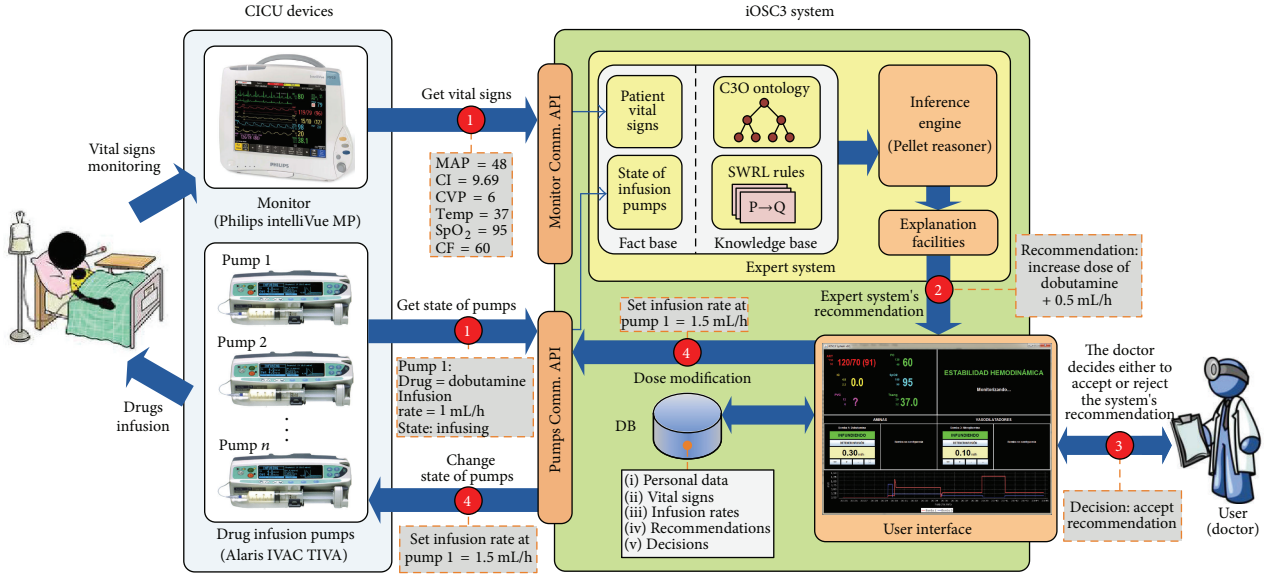


Figure 1: iOSC3 architecture and workflow.

in charge of analyzing the patient's condition and providing a recommendation about the treatment that should be applied to stabilize him/her. The expert system has 4 essential components: knowledge base, fact base, inference engine, and explanation facilities.

(i) The *knowledge base* contains all of the relevant domain-specific information, permitting the system to act as a specialized problem solver for the CICU domain. It has been constructed in direct collaboration with the physicians at the CICU of the Meixoeiro Hospital, who are the experts in the problem domain. It consists of two essential components: an ontology, which has been written in a formal language and represents the concepts used at the CICU (e.g., vital sign, medical device, infusion pump, drug, etc.) and their relations (e.g., cardiac index "is a" vital sign) and a group of IF-THEN rules that represent the protocol used by the CICU doctors to diagnose and treat patients. A detailed explanation about how the ontology and the rules were built will be provided in Section 3.

(ii) The *fact base* contains a collection of facts against which rule conditions are evaluated. It contains the values of the patient's vital signs and the state of each

infusion pump in a given time. These facts match the left sides of production rules to determine eligible rules for firing. As previously explained, these data are obtained from the CICU patient monitor and the infusion pumps by means of the Monitor and Pumps Communication APIs that have been developed.

(iii) The *inference engine* is the core of the expert system. It links the rules given in the knowledge base with the fact base and performs reasoning to reach a solution. As inference engine, we have decided to use Pellet [30], which is the leading choice for applications that need to reason about knowledge represented using OWL ontologies [9] and SWRL rules [31].

(iv) The *explanation facilities* enable the user to ask the expert system how a particular conclusion is reached. It has been shown that physicians are more likely to adhere to expert system recommendations when quality explanation facilities are available [32]. In fact, it has been reported that expert system advice is usually ignored when it is not accompanied by an explanation, even when users acknowledge its global good performance [33]. The iOSC3 system uses the standard explanation facilities provided by Pellet, which include explanation for SWRL rules.

The knowledge base and the facts base are loaded by the inference engine by means of Jena (http://jena.apache.org/), which is a Java framework for building Semantic Web applications that provides a complete collection of tools to manage ontologies. On the basis of this knowledge, the inference engine executes the reasoning process and provides the doctor with a recommendation about the patient's treatment (see step 2 in Figure 1).

2.1.4. Graphical User Interface. The user interface enables the communication between the doctor and the expert system. It has two main windows: the case management window, which allows the doctor to store and query the patient's general information, and the intelligent care window, which is the interface that the doctor uses to supervise and treat the patient.

The case management window (see Figure 2) allows the user to store and query information about the patients that are going to be treated by the system and configure the CICU devices for each one of them. This window has the following tabs.

(i) *Patient's Data.* The content of this tab is shown in Figure 2. It is aimed to save the patient's name, surname, clinical history number, bed number, age, height, and weight. It also allows the doctor to point out if the patient has some contraindications (cardiac, respiratory, surgical, or others) that may require different medical attention and may prevent the use of the system.

(ii) *Vital Signs Limits.* This tab allows the doctor to input the top and bottom limits for the patient's vital signs (mean arterial pressure, cardiac frequency, cardiac index, central venous pressure, temperature, and O2 saturation). These limits will be used by the expert system to determine the stability or instability of the patient and suggest appropriate treatment to ensure their recovery. As an example, the top and bottom limits for the mean arterial pressure could be 50 mmHg and 130 mmHg, respectively.

(iii) *Pumps Configuration.* The content of this tab is shown in Figure 3. It allows to set the number of pumps that are connected to the patient, as well as the specific type of drug in each pump and the initial infusion rate. The user also has to input the COM port number to which the pump is connected. As previously explained, current version of iOSC3 only allows the communication with infusion pumps that use the Alaris Pump Communications Protocol. However, due to its modular architecture, the system can be easily modified to allow the communication with other pump models. Regarding the number of pumps, the system allows to work with 2 pumps with amine drug and 2 pumps with vasodilator drug as maximum. This number of pumps, filled with the appropriate drugs, is enough to solve all the possible situations.

(iv) *Monitor Configuration.* The aim of this tab is to configure the connection to the patient monitor. Communication is achieved through LAN connection; so, it is necessary to provide the monitor IP (e.g., 169.254.127.255) and the communication port (e.g., 24105). Current version of the system is prepared to interact with Philips IntelliVue MP Series monitors.

(v) *Statistics.* This tab is used to show different statistics about the patient's treatment that can be useful to extract conclusions about the usefulness of the system. Examples of these statistics are patient's recovery time, total amount of drug supplied to each patient, and number of system's recommendations accepted and refused by the doctor.

When the user has entered all the data for a specific medical case, he/she can start the execution of the intelligent decision support system by clicking the "Start monitoring" button. Then, the intelligent care window is opened (see Figure 4). This window allows the doctor to monitor the patient's condition and supervise the treatment that is being supplied. It is divided into four sections, which display the following information.

(i) *Patient's Vital Signs.* The current values of the patient's vital signs, as well as the top and bottom limits that have been set for each parameter.

(ii) *State of Infusion Pumps.* The current state of the infusion pumps that are connected to the patient. For each pump, the screen shows the pump name and drug (e.g., name: pump1; drug: dobutamine), the infusion state (i.e., infusing or stopped), and the current infusion rate (e.g., 1.2 mL/h). It also provides buttons to start or stop the infusion and to modify the infusion rate.

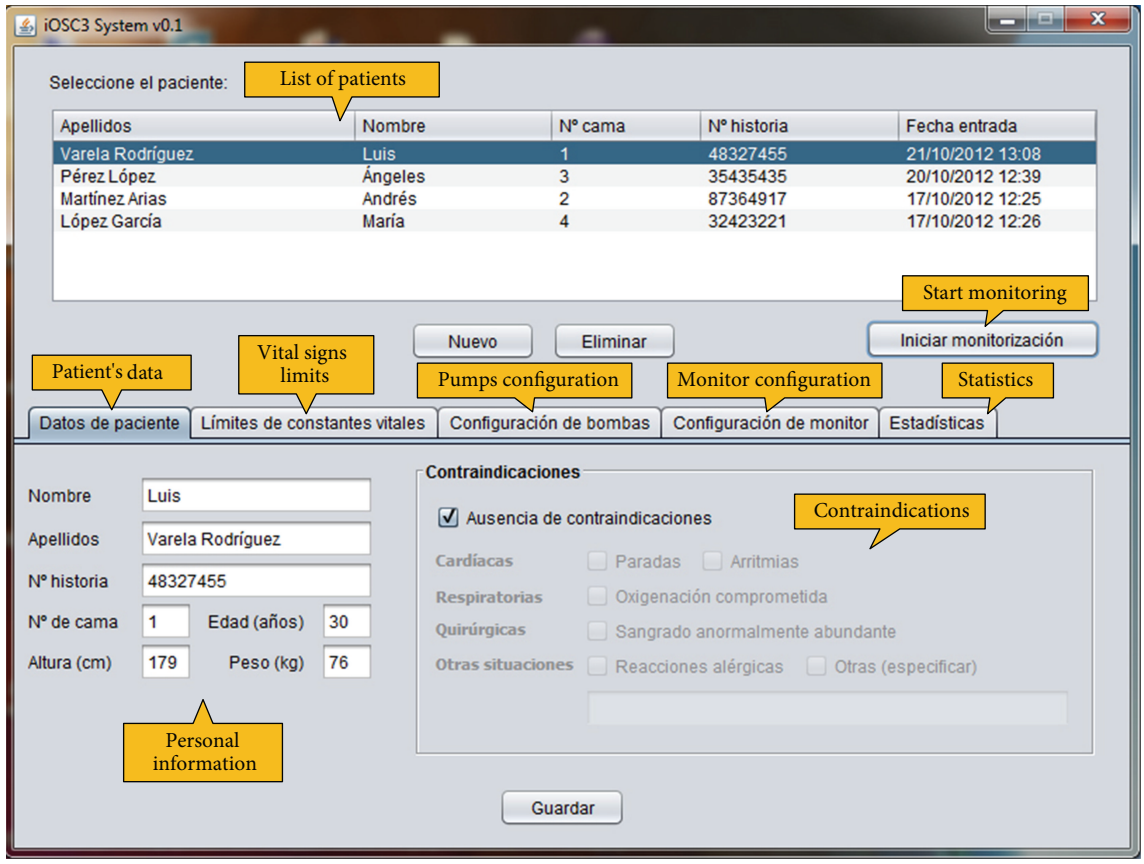


Figure 2: Case management window, showing the content of the patient's data tab.

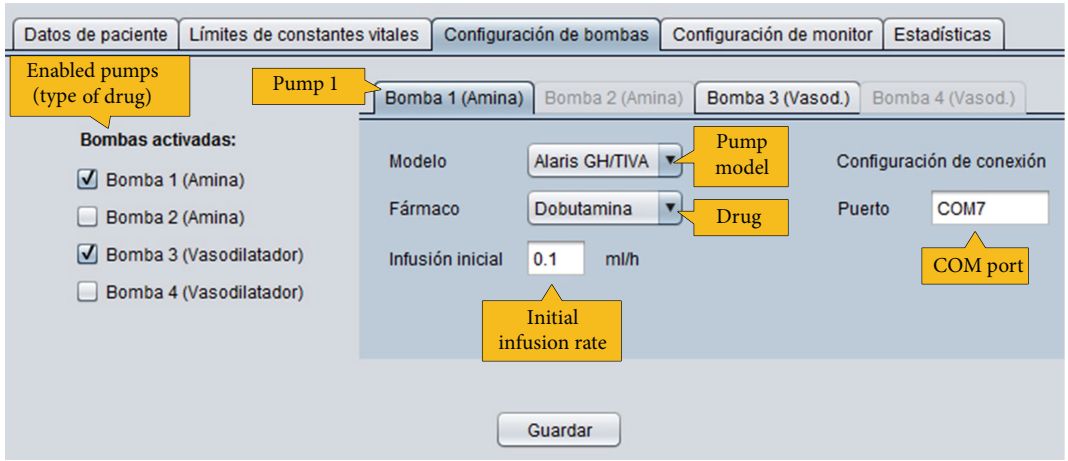


Figure 3: Pumps configuration tab.

(iii) *Evolution of Infusion Rates.* This region displays a graph that represents the infusion rate applied at each infusion pump during a certain period of time (e.g., last 20 minutes).

(iv) *Treatment Recommendation.* The expert system analyzes the patient's condition and the treatment that is being administered, and if the patient is in a situation of instability according to the expert knowledge contained in the knowledge base, the system provides a recommendation about the modifications that should be done on the infusion pumps to stabilize him/her. The doctor can accept or reject the system's recommendation. If the recommendation is accepted, the system automatically communicates with the infusion pumps to achieve the necessary modifications. The physician also can edit the infusion rate suggested by the system and set the infusion rate that

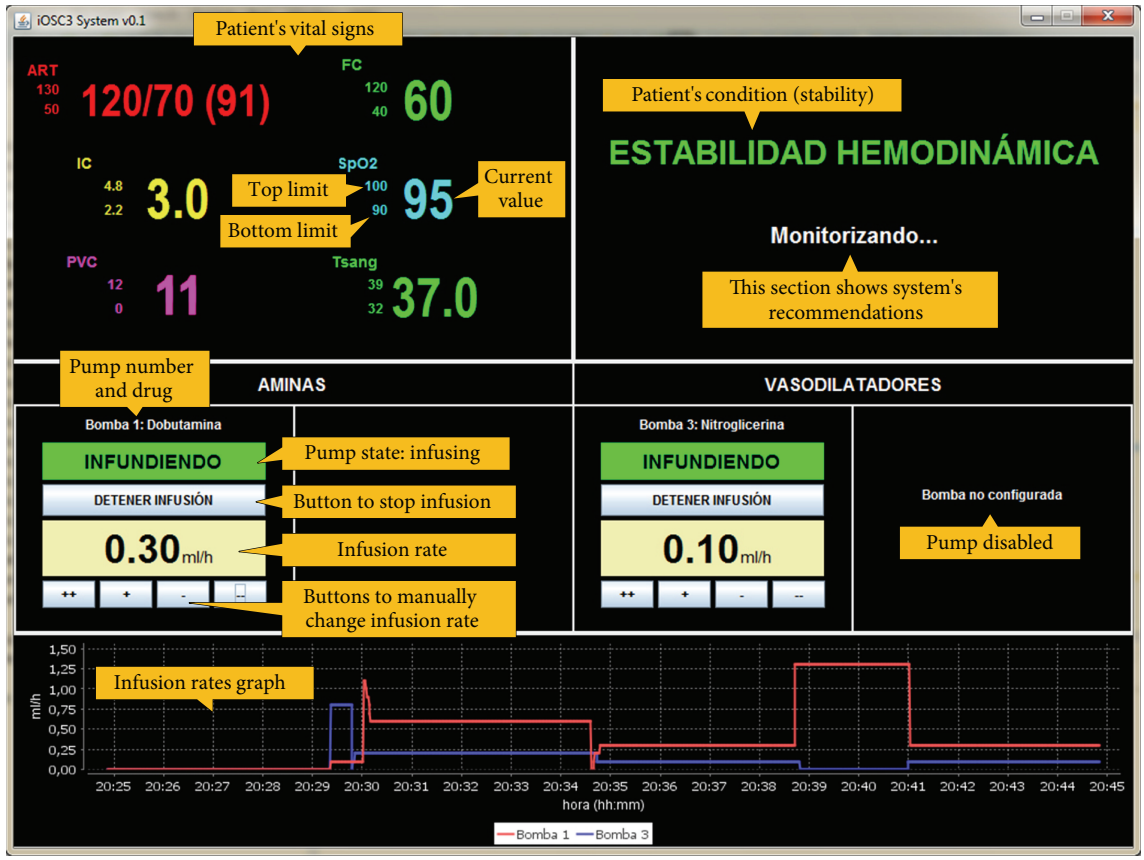


Figure 4: Intelligent care window.

Figure 5: iOSC3 installed in the CICU of the Meixoeiro Hospital.

he/she considers appropriate. If the doctor refuses the recommendation, the treatment is not modified and the system continues monitoring the patient.

2.1.5. Database. All the data managed by the iOSC3 system are stored into a MySQL database for subsequent analysis. For each patient, the system stores personal information, contraindications, limits of vital signs, pumps, and monitor configuration. In addition, while the patient is being monitored, the system periodically (each 5 seconds) stores his/her vital sign values, the state and infusion rates at each pump, the system's recommendations, and the decisions taken by the doctor. All these data are essential to extract conclusions about the usefulness of the system and to identify the aspects that should be improved.

Figure 5 shows the system installed in the CICU of the Meixoeiro Hospital. It is possible to see 2 Alaris infusion pumps, a Philips IntelliVue MP70 monitor, and a laptop that shows the system's intelligent care window.

3. Building the Critical Cardiac Care Ontology (C3O)

At this section, we explain the methods followed to extract the expert's knowledge and represent it formally as an OWL ontology. The ontology building process was guided by Methontology [34], one of the most popular methodologies for ontology development, following a bottom-up approach and based on the OBO Foundry principles (http://www.obofoundry.org/crit.shtml). The ontology development process was achieved according to the activities shown in Figure 6, which are described later.

3.1. Specification. The goal of this stage was to produce an informal ontology specification document written in natural language, that described general aspects of the ontology and its intended use. We wrote a document describing the ontology domain, purpose of the ontology, scenarios of use, intended users, level of formality, list of terms to be represented, and related bibliographic references. Initially, this document reflected many aspects that were not clear, but it has been gradually refined and expanded throughout the rest of the development process, giving rise to a structured text that reflects the ontology requirements in a concise and clear manner.

3.2. Knowledge Acquisition. This phase occurs largely in parallel with the rest of the ontology development process, especially with the specification stage, and decreases as the process moves forward. At this stage, we held a variety of meetings and interviews with the CICU doctors, who were asked to describe in detail the procedures they employ to monitor and treat patients. These interviews allowed to acquire a wide set of documented information (interview transcripts, books, scientific papers, diagrams, technical manuals, etc.) about the protocols followed in the CICU to treat patients, as well as regarding the technical details of the medical devices used at the CICU (patient monitor and drug infusion pumps).

3.3. Conceptualization. At this point, we structured the domain knowledge in a conceptual model. We identified the key terms (e.g., cardiac frequency, infusion pump, drug, etc.) and the relations between those terms (e.g., nitroglycerin "is a" vasodilator agent) and organized them in a taxonomic "is a" hierarchy following a top-bottom approach (see Figure 7). We also extracted the main rules that guide the decision making process and structured them as a list of IF-THEN rules. Table 2 provides a description of the most relevant classes and properties.

3.4. Implementation. This activity involved the representation of the ontology in a formal language. The ontology was formalized in OWL DL, a description logics-based sublanguage of the Ontology Web Language (OWL) [35]. It was chosen because it is highly expressive and it still retains computational completeness and decidability. In addition, several well-known reasoning systems are available for OWL DL, such as Pellet. The ontology was built using the Stanford University ontology editor Protégé (version 3.4.7) [36]. The inference rules were written in the Semantic Web Rule Language (SWRL) [31], which is the rule representation language recommended by the Semantic Web community and allows to express rules on the basis of ontology concepts. The rules were written using the SWRL Editor (see Figure 8), a development environment for working with SWRL rules in Protégé-OWL. When editing rules in this environment, users can directly refer to OWL classes, properties, and individuals within an OWL ontology. They also have direct access to a full set of built-ins described in the SWRL built-in specification and to all of the XML Schema data types [37]. The rules were stored as OWL individuals in the C3O ontology. Box 1 shows an example of rule used by the expert system written both in natural language and in the Semantic Web Rule Language.

The resulting ontology has been called Cardiac Critical Care Ontology (C3O). It contains 40 well-defined terms (classes) frequently used by experts in the area of CICUs organized as a taxonomy, 1 object property, 5 datatype properties, and a set of inference rules that guide the decision making process. The C3O ontology in OWL format is publicly available at http://tinyurl.com/cyeqq6x.

3.5. Integration. According to the knowledge-reuse principles proposed by the OBO Foundry [38], at this phase, we incorporated to the C3O ontology knowledge already provided by other ontologies. We checked if the identified concepts were already contained in other existing biomedical ontologies. Carrying out this process manually is a hard and time-consuming task; so, we used a biomedical ontology selection tool (the BIOSS system (http://bioss.ontologyselection.com/) [39, 40]). We observed that most of the concepts were distributed across different ontologies. Also, some concepts had not been previously defined. As an example, the MeSH ontology (version 2009_02_13) contains the concepts "dobutamine" and "infusion pump," but it does not contain the concept "Mean Arterial Pressure," which is contained in the NCI Thesaurus ontology (version 2008_05D). We referenced the concepts contained in other ontologies and created the concepts that had not been previously defined.

3.6. Evaluation. This task has been achieved during each phase and between phases. The term evaluation subsumes the terms verification and validation. As explained by Fernandez et al., verification refers to the technical process that guarantees the correctness of an ontology with respect to a frame of reference. Validation guarantees that the ontology corresponds to the system that it is supposed to represent [41]. We reviewed the ontology and the set of rules with respect to the requirements specification document in order to detect and solve incompleteness, inconsistencies, and redundancies.

3.7. Maintenance. Ontologies applied in real-world settings are continuously evolving [42]. When the requirements change, it is possible that some ontological knowledge is no more relevant to the application domain. In that case, this knowledge has to be removed from the ontology. In a similar way, sometimes it is necessary to add to the ontology new terms or relations that are necessary to support new application requirements.

4. Evaluation and Results

Evaluation of decision support systems is a global term that comprises two main stages: evaluation of the intrinsic properties of the system (technical evaluation) and evaluation of its actual use and utility (user's evaluation or assessment). Technical evaluation is divided in two tasks: verification and validation. Verification and validation assess the consistency, correctness, and completeness of the knowledge within the

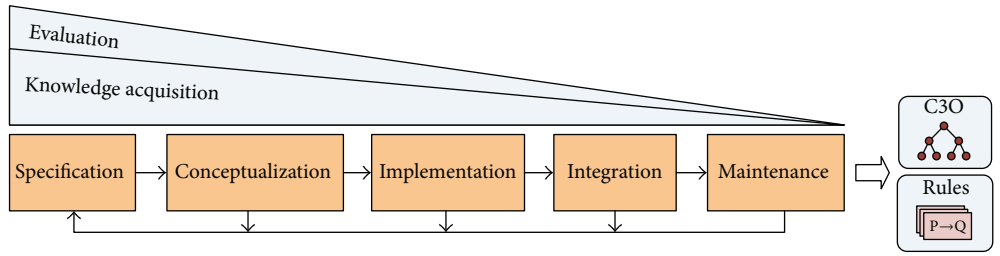


Figure 6: Ontology development process (a simplified version of the Methontology lifecycle [46]).

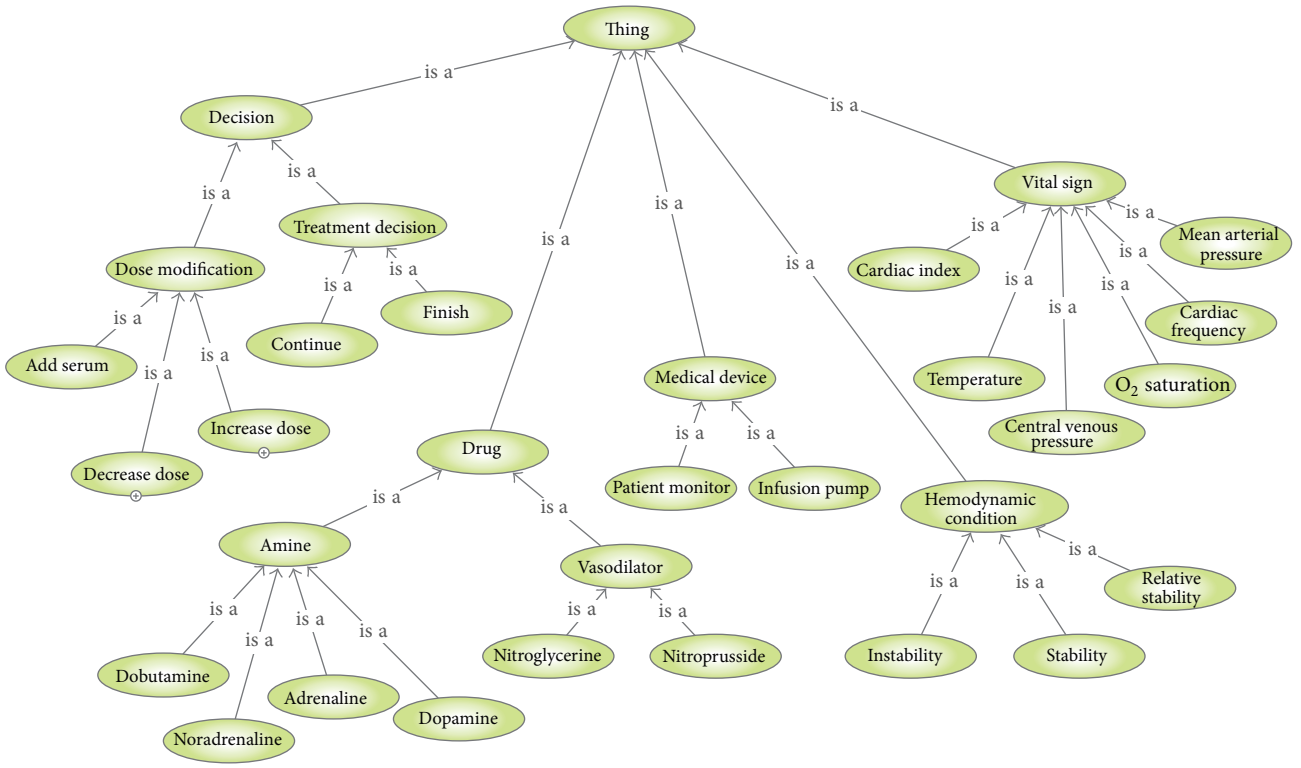


Figure 7: Fragment of the C3O hierarchy of classes.

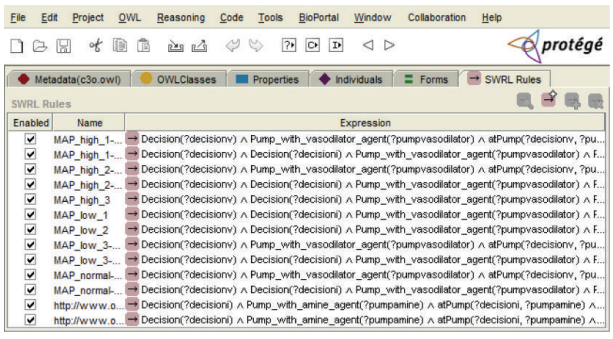


Figure 8: Screenshot of the SWRL Editor showing some of the C3O rules.

expert system, the quality of the solutions provided by the system, and the ability of the system to produce the same results given the same inputs [43].

In this section, we describe how the expert system was verified and validated. User's evaluation reflects the acceptance of the system by the end users and its performance in the field, and it will be suggested as a future work. It will require to design a clinical study to test how the system affects the structure, process, and outcome of health care encounters.

Verification of an expert system refers to "building the system right," that is, checking that the system is a correct implementation of the specification. This stage was carried out by the development team according to the requirements extracted from the medical experts. One of the major tasks in verifying expert systems is the verification of the knowledge contained within the knowledge base. This type of verification assesses the accuracy of the knowledge, because inaccuracy in the knowledge base results in inaccuracy for the whole expert system. As explained in Section 3.6, this kind of verification was achieved during the ontology development process.

Table 2: Main classes and properties contained in the C3O ontology.

Name	Type	Description
Decision	Class	A class that represents the different judgments made at the CICU to treat a patient.
Dose_modification	Class	Adjustment of the amount of drugs (infusion rate) that is being supplied to the patient.
Treatment_decision	Class	Choice about the treatment that is being administered to the patient.
Drug	Class	Substance used at the CICU to treat patients.
Amine	Class	Drug with vasoconstriction properties. Amines increase blood pressure, which raises organ perfusion pressure and preserves distribution of cardiac output to the organs. Examples: dobutamine, noradrenaline, and adrenaline.
Vasodilator	Class	Drug that relax the smooth muscle in blood vessels, which causes the vessels to dilate and decreases blood pressure. Examples: nitroglycerine and nitroprusside.
Medical_device	Class	Equipment used at the CICU to monitor or treat patients. Examples: patient monitor and infusion pump.
Hemodynamic_condition	Class	State of health of the patient with respect to the situation of the forces the heart has to develop to circulate blood through the cardiovascular system.
Vital_sign	Class	Indicator of a patient's general physical condition. Examples: cardiac index, temperature, and mean arterial pressure.
atPump	Object property	A property which allows to represent the dose modification that has to be applied at a specific infusion pump (domain: dose_modification; range: infusion_pump).
hasValue	Datatype property	A property which represents the values and limits of the patient's vital signs by means of its children: hasExactValue, hasLowerLimitValue, hasNormalValue, and hasUpperLimitValue.

Natural Language
"IF there are two infusion pumps connected to the patient, one containing a vasodilator drug and the other containing an amine drug, AND the Mean Arterial Pressure (MAP) is higher than its upper limit AND the Cardiac Index (Cl) is lower than the reference value (e.g. 2.2) AND the infusion rate at the vasodilator pump is lower than the maximum infusion rate; THEN the infusion rate at the vasodilator pump must be gradually increased"
Semantic Web Rule Language (SWRL)
`Decision (?decisionv) ∧ Decision (?decisioni) ∧ Pump_with_vasodilator_agent (?pumpvasodilator)`
`∧ Pump_with_amine_agent (?pumpamine) ∧ atPump (?decisionv, ?pumpvasodilator) ∧`
`atPump (?decisioni, ?pumpamine) ∧ Mean_Arterial_Pressure (?map) ∧`
`hasExactValue (?map, ?mapvalue) ∧ hasUpperLimitValue (?map, ?mapupperlimit) ∧`
`swrlb: greaterThan (?mapvalue, ?mapupperlimit) ∧ Cardiac_Index (?ci) ∧`
`hasExactValue (?ci, ?civalue) ∧ hasNormalValue (?ci, ?cinormalvalue) ∧`
`swrlb: lessThan (?civalue, ?cinormalvalue) ∧`
`hasUpperLimitValue (?pumpvasodilator, ?pumpvasodilatorlowerlimit) ∧`
`hasExactValue (?pumpvasodilator, ?pumpvasodilatorvalue) ∧`
`swrlb: lessThan (?pumpvasodilatorvalue, ?pumpvasodilatorlowerlimit)`
`→ Increase_dose_maxlimit_gradually (?decisionv)`

Box 1: Example of rule written in natural language and SWRL.

Table 3: Example of test case. The MAP value (40.0) is lower than its lower limit (50.0). Other patient parameters have normal values. In this situation, the decision would be to decrease the vasodilator infusion rate and increase the amine infusion rate.

	Parameter	Unit	Value	Lower limit	Upper limit
Vital parameters	Mean arterial pressure (MAP)	mmHg	40.0	50.0	90.0
	Oxygen saturation (SpO_2)	%	92.0	90.0	100.0
	Central venous pressure (CVP)	mmHg	10.0	4.0	20.0
	Cardiac frequency (CF)	bpm	76.0	40.0	120.0
	Cardiac index (CI)	$L/min/m^2$	2.3	Ref. value: 2.2	
	Temp (T)	°C	37.1	32.0	39.0
Infusion rates	Vasodilator pump	mL/h	4.0	0.0	10.0
	Amine pump	mL/h	2.0	0.0	10.0
Expected recommendation	Decrease the infusion rate at the vasodilator pump and increase the infusion rate at the amine pump until the MAP reaches normal values				

Table 4: Summary of validation results.

Parameter	Value
Number of test cases	14
Correct decisions	14
Incorrect decisions	0
Precision	100%

Validation generally is regarded as a more complex task than verification. It refers to "building the right system," that is, substantiating that it performs with an acceptable level of accuracy the real-world tasks for which it was intended. Validation is the cornerstone of evaluation since, for example, a highly efficient implementation of an invalid system is useless [44]. There are different ways to validate an expert system. We have decided to achieve a validation based on test cases, because this method has been reported to be the dominant strategy for the systemic validation of expert systems [45].

Preparing a high-quality set of test cases is crucial to achieve an accurate and complete validation. This set must offer an adequate domain coverage. A good criterion to construct the set of test cases is covering all the possible system outputs that could be generated in a particular moment [19]. We defined a set of 14 domain-representative test cases. Then, we executed the system and compared the system's output with the expected output, provided by a medical expert. Each test case consisted of a set of input parameters (values of patient's vital signs and pump infusion rates) and an expected output (system's recommendation). The test cases were created on the basis of real data, collected at the CICU and inserted into the system by means of a software application. An example of one of these test cases is shown in Table 3. After executing all test cases, the system was able to achieve an overall precision of 100% (see Table 4).

In spite of that a comprehensive user's evaluation is suggested as a future work, during the development of the system, we have received some remarks from the physicians at the CICU of the Meixoeiro Hospital that may be considered of interest. The feedback received is summarized as follows.

The medical experts who have interacted with the system or discussed about it during the development process consider that it can be a useful support tool for their daily work, especially in situations of high workload or uncertainty. The system is able to provide a recommendation about the patient's treatment in a matter of a few seconds, and physicians think that, even if the system is not completely accurate, it can be extremely useful to guide them to the proper decision. Other comment that we have received is that physicians are very comfortable with the system's user interface. The interface has been designed following the look and feel (i.e., colors, screen distribution, fonts, etc.) of existing patient monitors; so, it looks familiar to the medical experts, and they learn how to use it quickly. Finally, another feature of iOSC3 that is considered very valuable by physicians is that all the information about the status of the patient and the infusion pumps is stored into a database for further analysis, regardless of whether the system is used for decision support or not. After a patient leaves the CICU, physicians can use all these data to study his/her evolution and discover human errors that have occurred, identify aspects of medical protocols that could be improved, compare the effectiveness of different drugs, and so forth.

5. Conclusions and Future Research

Patient monitoring and management at the ICU is a difficult task involving acquisition and processing of a huge volume of complex data. ICU physicians have to analyze and interpret these data to make quick decisions, which frequently imply modifications on the dosage of drugs being administered. However, the human factor is sometimes a source of mistakes that lead to inaccurate or erroneous decisions about patients' care. The development of decision support solutions that help physicians to manage and process the continuous flow of information and to make quick and reliable decisions about the treatment of patients can provide multiple benefits, including improved quality of care and an overall reduction in cost. In general, experts in critical environments are very interested in the development of such kind of systems because they make their daily work easier and help them to avoid mistakes in patients' treatment.

In this work, an ontology-based system for intelligent supervision and treatment of critical patients with acute cardiac disorders has been presented. The system is based on expert knowledge, which has been formalized in the form of an ontology and a set of semantic rules. This ontology contains the main concepts used by experts in CICUs, the relationships between these concepts, and the set of inference rules that guide the decision making process. To the best of our knowledge, this is the first time that knowledge to treat critical cardiac patients has been formally represented as an ontology and used as the basis to build a decision support system.

Future research is mainly focused on designing a study to be achieved at the CICU of the Meixoeiro Hospital that will allow to evaluate the use of the system in clinical practice and how it affects the treatment and recovery of patients.

Acknowledgments

This work is supported by the Carlos III Health Institute (Grant FIS-PI10/02180), the Ibero NBIC Network (Ref. 209RT0366) funded by CYTED, and Grants CN2012/217 (REGICC), CN2011/034 (Programa de consolidación y estructuración de unidades de investigación competitivas), and CN2012/211 (Agrupación estratégica) from the Xunta de Galicia. It is also cofunded by FEDER (European Union). The authors thank the staff of the Service of Anesthesiology, Resuscitation and Intensive Care of the Meixoeiro Hospital for their kind help and valuable collaboration. They also would like to thank Chandana Samaranayake for his support during the development of the Monitor Communication API.

References

[1] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3465–3469, 2009.

[2] Y. Donchin, D. Gopher, M. Olin et al., "A look into the nature and causes of human errors in the intensive care unit," *Critical Care Medicine*, vol. 23, no. 2, pp. 294–300, 1995.

[3] P. J. Pronovost, D. A. Thompson, C. G. Holzmueller, L. H. Lubomski, and L. L. Morlock, "Defining and measuring patient safety," *Critical Care Clinics*, vol. 21, no. 1, pp. 1–19, 2005.

[4] E. Camiré, E. Moyen, and H. T. . Stelfox, "Medication errors in critical care: risk factors, prevention and disclosure," *Canadian Medical Association Journal*, vol. 180, pp. 936–943, 2009.

[5] E. Tissot, C. Cornette, P. Demoly, M. Jacquet, F. Barale, and G. Capellier, "Medication errors at the administration stage in an intensive care unit," *Intensive Care Medicine*, vol. 25, no. 4, pp. 353–359, 1999.

[6] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, 2001.

[7] T. R. Gruber, *The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases: Knowledge Systems Laboratory*, Computer Science Department Stanford University, 1991.

[8] O. Lassila and R. R. Swick, *Resource Description Framework (RDF) Model and Syntax Specification*, 1999.

[9] S. Bechhofer, F. van Harmelen, J. Hendler et al., "OWL web ontology language reference," W3C Recommendation, vol. 10, pp. 2006-01, 2004.

[10] Y. A. Lussier and O. Bodenreider, "Clinical ontologies for discovery applications," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. J. O. Baker and K.-H. Cheung, Eds., pp. 101–119, Springer, New York, NY, USA, 2007.

[11] A. C. Yu, "Methods in biomedical ontology," *Journal of Biomedical Informatics*, vol. 39, no. 3, pp. 252–266, 2006.

[12] N. F. Noy, N. H. Shah, P. L. Whetzel et al., "BioPortal: ontologies and integrated data resources at the click of a mouse," *Nucleic Acids Research*, vol. 37, no. 2, pp. W170–W173, 2009.

[13] M. A. Musen, Y. Shahar, H. Edward et al., "Clinical decision-support systems," *Biomedical Informatics*, pp. 698–736, 2006.

[14] F. T. De Dombal, D. J. Leaper, and J. C. Horrocks, "Human and computer aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians," *British Medical Journal*, vol. 1, no. 5904, pp. 376–380, 1974.

[15] R. A. Miller, H. E. Pople, and J. D. Myers, "Internist-I, an experimental computer-based diagnostic consultant for general internal medicine," *The New England Journal of Medicine*, vol. 307, no. 8, pp. 468–476, 1982.

[16] E. H. Shortliffe, *Computer-Based Medical Consultations: MYCIN*, Elsevier, New York, NY, USA, 1976.

[17] K. A. Kumar, Y. Singh, and S. Sanyal, "Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU," *Expert Systems with Applications*, vol. 36, no. 1, pp. 65–71, 2009.

[18] Y. Si, J. Gotman, A. Pasupathy, D. Flanagan, B. Rosenblatt, and R. Gottesman, "An expert system for EEG monitoring in the pediatric intensive care unit," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 6, pp. 488–500, 1998.

[19] Á. García-Crespo, A. Rodríguez, M. Mencke, J. M. Gómez-Berbís, and R. Colomo-Palacios, "ODDIN: ontology-driven differential diagnosis based on logical inference and probabilistic refinements," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2621–2628, 2010.

[20] H. Ying and L. C. Sheppard, "Regulating mean arterial pressure in postsurgical cardiac patients," *IEEE Engineering in Medicine and Biology Magazine*, vol. 13, no. 5, pp. 671–677, 1994.

[21] A. S. Nocedal, J. K. Gerrikagoitia, and I. Huerga, "Supporting clinical processes with semantic web technologies: a case in breast cancer treatment," *International Journal of Metadata, Semantics and Ontologies*, vol. 5, no. 4, pp. 309–320, 2010.

[22] J. M. Blum, G. H. Kruger, K. L. Sanders, J. Gutierrez, and A. L. Rosenberg, "Specificity improvement for network distributed physiologic alarms based on a simple deterministic reactive intelligent agent in the critical care environment," *Journal of Clinical Monitoring and Computing*, vol. 23, no. 1, pp. 21–30, 2009.

[23] R.-C. Chen, Y.-H. Huang, C.-T. Bau, and S.-M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," *Expert Systems with Applications*, vol. 39, no. 4, pp. 3995–4006, 2012.

[24] S. H. Liao, "Expert system methodologies and applications-a decade review from 1995 to 2004," *Expert Systems with Applications*, vol. 28, no. 1, pp. 93–103, 2005.

[25] S. Le Moigno, J. Charlet, D. Bourigault, P. Degoulet, and M. C. Jaulent, "Terminology extraction from text to build an ontology in surgical intensive care," *Proceedings of the AMIA Symposium*, pp. 430–434, 2002.

[26] N. F. De Keizer, A. Abu-Hanna, R. Cornet, J. H. M. Zwetsloot-Schonk, and C. P. Stoutenbeek, "Analysis and design of an ontology for intensive care diagnoses," *Methods of Information in Medicine*, vol. 38, no. 2, pp. 102–112, 1999.

[27] G. Booch, J. Rumbaugh, and I. Jacobsen, *The Unified Software Development Process—The Complete Guide to the Unified Process from the Original Designers*, 1999.

[28] Philips Medical Systems, *Philips Intellivue Patient Monitor X2, MP Series, MX Series. Data Export Interface Programming Guide*, 2008.

[29] Alaris Medical Systems, *Asena Syringe Pump Communications Protocol. Issue: 3.0*, 2003.

[30] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: a practical OWL-DL reasoner," *Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007.

[31] I. Horrocks, P. F. Patel-Schneider, H. Boley et al., "SWRL: a semantic web rule language combining OWL and RuleML," W3C Member submission, vol. 21, p. 79, 2004.

[32] K. W. Darlington, "Designing for explanation in health care applications of expert systems," *SAGE Open*, vol. 1, 2011.

[33] M. S. Gönül, D. Önkal, and M. Lawrence, "The effects of structural characteristics of explanations on use of a DSS," *Decision Support Systems*, vol. 42, no. 3, pp. 1481–1493, 2006.

[34] M. F. López, A. Gómez-Pérez, J. P. Sierra, and A. P. Sierra, "Building a chemical ontology using methontology and the ontology design environment," *IEEE Intelligent Systems and Their Applications*, vol. 14, no. 1, pp. 37–46, 1999.

[35] L. W. Lacy, *Owl: Representing Information Using the Web Ontology Language*, Trafford Publishing, 2005.

[36] N. F. Noy, M. Sintek, S. Decker, M. Crubézy, R. W. Fergerson, and M. A. Musen, "Creating semantic web contents with protege-2000," *IEEE Intelligent Systems and Their Applications*, vol. 16, no. 2, pp. 60–71, 2001.

[37] M. O'Connor, S. W. Tu, C. I. Nyulas et al., "Querying the semantic web with SWRL," *Advances in Rule Interchange and Applications*, pp. 155–159, 2007.

[38] B. Smith, M. Ashburner, C. Rosse et al., "The OBO Foundry: coordinated evolution of ontologies to support biomedical data

integration," *Nature Biotechnology*, vol. 25, no. 11, pp. 1251–1255, 2007.

[39] M. Martínez-Romero, J. M. Va'zquez-Naya, J. Pereira et al., "A multi-criteria approach for automatic ontology recommendation using collective knowledge," in *Recommender Systems for the Social Web*, pp. 89–104, Springer, 2012.

[40] M. Martínez-Romero, J. M. Va'zquez-Naya, C. R. Munteanu et al., "An approach for the automatic recommendation of ontologies using collaborative knowledge," in *Lecture Notes in Artificial Intelligence*, vol. 6277, pp. 74–81, 2010.

[41] M. Fernandez, A. Gomez-Perez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," in *Proceedings of the Spring Symposium Series on Ontological Engineering (AAAI '97)*, pp. 33–40, 1997.

[42] A. Maedche and R. Volz, "The ontology extraction & maintenance framework Text-To-Onto," in *Proceedings of the Workshop on Integrating Data Mining and Knowledge Management*, 2001.

[43] S. Smith and A. Kandel, *Verification and Validation of Rule-Based Expert Systems*, CRC, 1993.

[44] R. M. O'Keefe, *Validation of Expert System Performance*, Department of Computer Science, Virginia Polytechnic Institute and State University, 1986.

[45] R. M. O'Keefe and D. E. O'Leary, "Performing and managing expert system validation," *Advances in Expert Systems for Management*, vol. 1, p. 141, 1993.

[46] O. Corcho, M. Fernández-López, and A. Gómez-Pérez, "Methodologies, tools and languages for building ontologies. Where is their meeting point?" *Data and Knowledge Engineering*, vol. 46, no. 1, pp. 41–64, 2003.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 548208, 5 pages
doi:10.1155/2012/548208

Research Article

Locomotor Development Prediction Based on Statistical Model Parameters Identification

A. V. Wildemann,[1] A. A. Tashkinov,[1] and V. A. Bronnikov[1,2]

[1] *The Center of Mathematical Modelling of Medicosocial Systems and Processes, The State National Research Politechnical University of Perm, Perm 614990, Russia*
[2] *The Department of Physical Culture and Health, The Perm State Medical Academy, Perm 614000, Russia*

Correspondence should be addressed to A. V. Wildemann, wildemann@mail.ru

Received 2 July 2012; Revised 8 October 2012; Accepted 23 October 2012

Academic Editor: Alejandro Rodríguez-González

This paper introduces an approach for parameters identification of a statistical predicting model with the use of the available individual data. Unknown parameters are separated into two groups: the ones specifying the average trend over large set of individuals and the ones describing the details of a concrete person. In order to calculate the vector of unknown parameters, a multidimensional constrained optimization problem is solved minimizing the discrepancy between real data and the model prediction over the set of feasible solutions. Both the individual retrospective data and factors influencing the individual dynamics are taken into account. The application of the method for predicting the movement of a patient with congenital motility disorders is considered.

1. Introduction

Multidimensional statistics methods, including predicting and dependencies analysis, are required to research complex systems with many random factors [1–4]. In order to obtain more reliable individual prediction, the patient specific information needs to be taken into account to correct statistical model parameters.

Modeling of biomedical systems provides actual application of mathematical methods and algorithms. This paper studies the problem of locomotor development prediction for people with congenital motility abnormalities. The authors propose an approach to identify the parameters of a statistical predictive model taking into account the available set of individual data.

2. Current State

Mathematical methods and information technologies are important tools in modern biomedical research. Together with physicomechanical modeling of physiological systems [5–7], the study of statistical properties of disease courses for patients with similar diagnoses is also of certain interest. Statistical predictive modeling in medicine covers a wide range of areas including cardiology [8–10], pulmonology [11–13], neurology [14, 15], and others.

One of the disadvantages of statistical predictive models is that they are only applicable to an average individual. In order to get a more accurate prediction it is necessary to identify statistical predictive model parameters using data for a concrete person. Using this approach which is known as "individual prediction" method [16, 17] in medical science is of major interest for the modern interdisciplinary scientific development.

3. Materials and Methods

3.1. Initial Data Analysis. The motility index Y quantifies the level of locomotor development. In order to calculate Y an expert evaluates 12 different groups of locomotor skills. The expert estimations are then arranged according to the five-point scale and the index Y is defined as the sum of the estimations. Due to the construction algorithm the index Y is defined within the interval $[0, 60]$. The best locomotor

development corresponds to the top boundary of parameter Y and the worst locomotor development corresponds to the lowest value.

The motility index dependence on age for a group of patients is considered to be a random process $Y(t)$, where t is the patient's age. The process $y^{(J)}(t)$ corresponds to each patient J in the group.

To analyze the stochastic process structure the following correlation function is used as follows:

$$r_y(t,t') = \frac{M\left[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')\right]}{\sqrt{D_y(t)D_y(t')}}, \tag{1}$$

where $\overset{\circ}{Y}(t)$ and $\overset{\circ}{Y}(t')$ are time slices of the centred random process $\overset{\circ}{Y}(t)$ at the time t and t', $D_y(t)$ and $D_y(t')$ are dispersions of $Y(t)$ at the moments t and t', and M is the expectation operator.

The available statistics database contains 157 observations of cerebral palsy patients. The research sample was based on occasional patient visits to a rehabilitation centre during their first nine years of life. Experts determined each patient's motility index during their visits. Detailed information was collected on each patient: medical status of parents and close relatives prior to the individual's conception and characteristics of prenatal and perinatal periods (prenatal and intranatal factors).

Despite the presence of relatively large number of observations, it was only possible to focus on 5 people who were constantly observed during the entire research period. Given the data limitation the study of locomotor skills development was conducted on a group of 5 people ($N = 5$) and the complete initial sample of 157 observations was used for calculating the average parameters.

A scatter diagram of random process $Y(t)$ is shown in Figure 1, where t is the age (in months). Figure 2 shows the graph of the corresponding correlation function for fixed age $t = 15$ (months).

When the correlation function is close to one, the process is characterized by a strong dependence between the time slices. This indicates that the process realizations are similar. The similarity is the necessary condition for the application of the proposed method.

3.2. General Algorithm for Individual Prediction. Individual prediction of motility index is based on the following algorithm.

The first step determines the age dependence of the average motility. The consecutive selection method of the exponential terms [16] is used for this purpose. According to this method, the average motility index trend is represented as follows:

$$f(t) = a_0 + \sum_{q=1}^{Q} a_q\left(1 - e^{-\lambda_q t}\right), \tag{2}$$

which includes the following unknown parameters: a_0, a_q, $\lambda_q (q = \overline{1,Q})$.

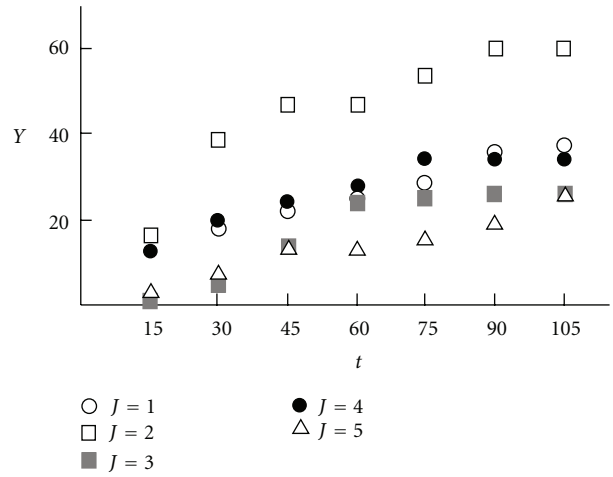


Figure 1: The realization of motility index.

The individual motility index trend is constructed as bounded from above monotonically increasing function:

$$F^{(J)}(t) = A_0^{(J)} + \sum_{q=1}^{Q} A_q^{(J)}\left(1 - e^{-\lambda_q t}\right). \tag{3}$$

This paper studies the locomotor skills accumulation process only. The loss of such skills is not taken into account therefore the monotonically increasing assumption could be used. The motility index boundaries are defined by their calculation method (from 0 to 60).

The generalized factor of prenatal and intranatal conditions is introduced as follows:

$$\Phi = \sum_{r} \delta_r P_r, \tag{4}$$

where P_r is the risk factors during pregnancy and birth that have the highest influence on the motility index, and δ_r is the correlation between P_r and the motility index.

The assumption of P_r factors being responsible for development delay and, as such, affecting the absolute term value $A_0^{(J)}$ is used.

The motility index value $A_0^{(J)}|_{\Phi=0}$ at time 0 with no prenatal and intranatal risk factors taken into account is calculated by solving the following convex optimization problem with constraints in the form of inequalities.

Find $\widetilde{A}_0^{(J)}|_{\Phi=0}$ and $\widetilde{A}_q^{(J)} (q = \overline{1,Q})$ such that:

$$\sum_{\gamma=1}^{\Theta_0}\left(A_0^{(J)}\Big|_{\Phi=0} + \sum_{q=1}^{Q} A_q^{(J)}\left(1 - e^{-\lambda_q t_\gamma}\right) - y_\gamma^{(J)}\right)^2 \longrightarrow \min,$$

$$A_q^{(J)} \geq 0\left(q = \overline{1,Q}\right), \tag{5}$$

$$A_0^{(J)}\Big|_{\Phi=0} + \sum_{q=1}^{Q} A_q^{(J)}\left(1 - e^{-\lambda_q t_\Theta}\right) \leq y_{\max},$$

where $\lambda_q (q = \overline{1,Q})$ are known values calculated during the average trend $f(t)$ identification, Θ_0 is the number of real

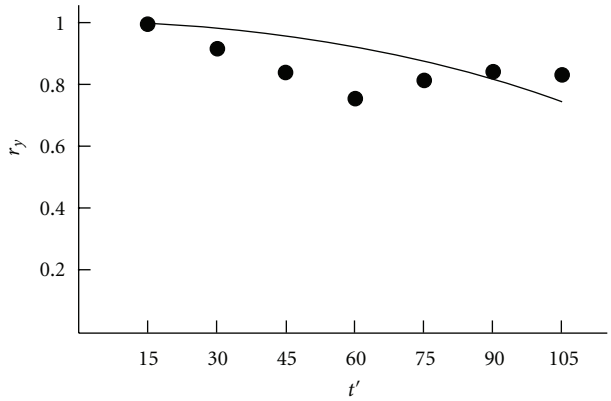


Figure 2: The correlation function of motility index.

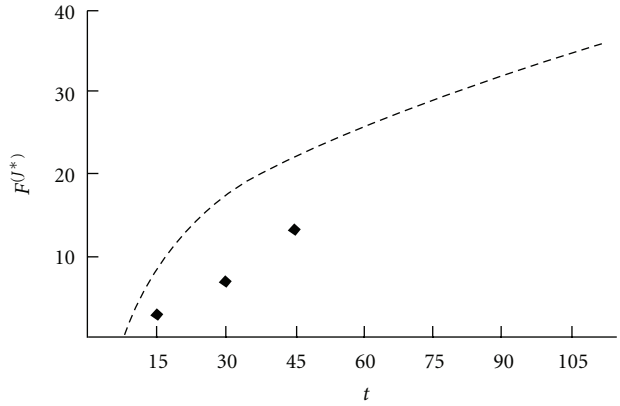


Figure 3: The individual data (black dots) and the average trend (dashed line) of motility index.

motility index values in the interval of the initial individual observation τ_0, $y_\gamma^{(J)}$ is the J's patient real motility index value at the moment of time $t_\gamma \in \tau_0$, t_Θ is the prediction horizon, and $y_{\max}$ is the possible highest value of motility index.

The first expression in system (5) is the minimum condition for the sum of square deviations from real motility index values. The second expression is the motility monotonic growth condition. The third one is the motility index restriction condition.

The problem (5) could either be solved numerically or analytically via the Kuhn-Tucker theorem [18].

The angular coefficient μ (which determines how strong the P_r factors influence the initial motility index changes) is calculated using linear approximation of the real motility index dependence at the initial time on the generalized prenatal and intranatal factors for a group of patients. The calculated coefficient μ is used for $A_0^{(J)}$ adjustment taking the influence of prenatal and intranatal factors into account: $A_0^{(J)} = A_0^{(J)}|_{\Phi=0} + \mu\Phi^{(J)}$. The other individual coefficients $A_q^{(J)}(q = \overline{1,Q})$ are calculated by solving the optimization problem (5) with the adjusted coefficient $A_0^{(J)}$ obtained during the previous step.

4. Results and Discussion

The practical applicability of this method is illustrated for a patient J^*. The person J^* is a cerebral palsy patient and was observed at the Perm Center of Complex Rehabilitation for People with Disabilities during the first nine years of life.

The entire observation time period is divided into two intervals: base period (first four years) and prediction period (next five years). The first period data is used to develop the motility index prediction model while the second period data serve as a test for prediction results (Table 1).

The initial locomotor development dynamics of the observed patient (black dots in Figure 3) differ significantly from the age dependence of average motility index for a group of patients with similar diagnosis (dotted line in Figure 3). Therefore the prediction based on the average trend only will not produce acceptable accuracy in final results. At the same time the prediction based on the individual data only will also not be accurate since the base period is shorter than the prediction period. The proposed method which corrects the average trend parameters using individual patient data is further carried out to increase the prediction accuracy.

Table 1: The individual data of motility index.

	Prediction base period			Prediction period			
t_γ, month	15	30	45	60	75	90	105
$y_\gamma^{(J^*)}$	3.00	7.00	13.25	13.25	15.50	19.25	25.00

Using the method of sequential selection of exponential terms [16] results in the expression for the average motility index trend:

$$f(t) = -16.49 + 66.87\left(1 - e^{-0.01t}\right) + 28.39\left(1 - e^{-0.09t}\right). \quad (6)$$

The correlation analysis method is applied to identify a set of variables that are further used as prenatal and intranatal conditions (Table 2) that significantly influence the motility index for a group of patients with cerebral palsy. The sample of 157 observations was used to calculate the correlation coefficient [17].

The individual values of prenatal and intranatal factors are shown in Table 3: "0" means "no", "1" means "yes".

The generalized factor of prenatal and intranatal conditions with the individual data and the correlations shown in Tables 2 and 3 being taken into account has the value $\Phi^{(J^*)} = 0.67$.

Solving the system (5) gives the motility index at the initial time in the absence of prenatal and intranatal risk factors: $A_0^{(J^*)}|_{\Phi=0} = -3.00$.

The coefficient μ obtained by the least-squares method [19] for the group of patients is -16.68. The corresponding adjusted initial motility index is calculated as: $A_0^{(J^*)} = A_0^{(J^*)}|_{\Phi=0} + \mu\Phi^{(J^*)} = -14.17$. The remaining coefficients are $A_1^{(J^*)} = 60.52$, $A_2^{(J^*)} = 17.04$.

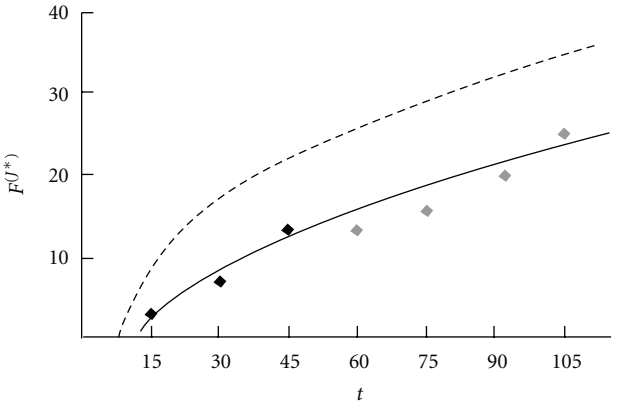


Figure 4: The control data (gray dots) and individual prediction (solid line) of motility index.

Table 2: The set of most significant prenatal and intranatal factors.

P_r	The factors	δ_r
P_1	Signs of intrauterine infection	−0.50
P_2	Signs of fetal hypoxia	−0.39
P_3	Extremely low weight at birth	−0.28

Table 3: The individual data of prenatal and intranatal factors.

$P_1^{(J^*)}$	$P_2^{(J^*)}$	$P_3^{(J^*)}$
0	1	1

The identified individual motility index trend $F^{(J^*)}(t) = -14.17 + 60.52(1 - e^{-0.01t}) + 17.04(1 - e^{-0.09t})$ is shown in Figure 4 (solid line). Compared to the average trend the individual predictive model is much closer to the control data (gray dots). The determination coefficient is $R^2 = 0.92$.

5. Conclusions

The proposed method of parameters identification for a statistical predictive model based on data set of patient information can be successfully applied conditioned on the correlation function being close to one. This method is based on multidimensional statistical analyses [1–4] and individual prediction approach [16, 17].

The application of the method to statistical modeling of medical and social systems resulted in an algorithm which does not require modeling of physical and mechanical processes. The obtained algorithm is based on statistical analysis only. The application of the algorithm significantly improves the prediction accuracy compared to the predictions based on average statistics. The improvement has been successfully demonstrated in an example of individual locomotor development prediction for a patient with congenital motility disorders.

The results show the advantages of the proposed method when the predicted index has a significant variation within the group with the prediction interval being larger than the base period.

Acknowledgment

This work was supported by the Russian Foundation for Basic Research (Grant 10-04-96096-r_ural_a).

References

[1] J. Hartung, B. Elpelt, and K. H. Klosener, *Statistik*, Oldenbourg, Munchen, Germany, 1993.
[2] C. J. Huberty, "Discriminant analysis," *Review of Educational Research*, vol. 45, pp. 543–598, 1975.
[3] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Pearson, 6th edition, 2007.
[4] D. A. Powers and Y. Xie, *Statistical Methods for Categorical Data Analysis*, Academic Press, New York, NY, USA, 2000.
[5] Y. C. Fung, *Biomechanics. Mechanical Properties of Living Tissues*, Springer, New York, NY, USA, 2nd edition, 1993.
[6] A. G. Kuchumov, Y. I. Nyashin et al., "Biomechanical approach to biliary system modelling as a step towards virtual physiological human project," *Russian Journal of Biomechanics*, vol. 15, no. 2, pp. 28–42, 2011.
[7] J. Nagatomi, *Mechanobiology, Handbook*, CRC Press Taylor & Fransis Group, 2011.
[8] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, Calif, USA, 1984.
[9] A. A. Peel, T. Semple, I. Wang, W. M. Lancaster, and J. L. Dall, "A coronary prognostic index for grading the severity of infarction," *British Heart Journal*, vol. 24, no. 6, pp. 745–760, 1962.
[10] I. G. Rusyak, L. A. Leshchinskii, A. F. Farkhutdinov, I. V. Logacheva, S. B. Ponomarev, and V. G. Sufiyanov, "Mathematical model for predicting the recovery of working ability in patients with myocardial infarction," *Biomedical Engineering*, vol. 33, no. 4, pp. 167–171, 1999.
[11] T. Austin, S. Iliffe, M. Leaning, and M. Modell, "A prototype computer decision support system for the management of asthma," *Journal of Medical Systems*, vol. 20, no. 1, pp. 45–55, 1996.
[12] O. N. Keene, P. M. A. Calverley, P. W. Jones, J. Vestbo, and J. A. Anderson, "Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited," *European Respiratory Journal*, vol. 32, no. 1, pp. 17–24, 2008.
[13] T. A. Lieu, C. P. Quesenberry, M. E. Sorel, G. R. Mendoza, and A. B. Leong, "Computer-based models to identify high-risk children with asthma," *American Journal of Respiratory and Critical Care Medicine*, vol. 157, no. 4, pp. 1173–1180, 1998.
[14] O. Yu. Rebrova and O. A. Ishanov, "Development and realization of the artificial neural network for diagnosing different types of stroke," *Optical Memory & Neural Networks*, vol. 13, no. 3, pp. 159–162, 2004.
[15] O. Rebrova, V. Kilikowski, S. Olimpieva, and O. Ishanov, "Expert system and neural network for stroke diagnosis," *International Journal of Information Technology and Intelligent Computing*, vol. 1, no. 2, pp. 441–453, 2006.
[16] A. A. Tashkinov, A. V. Wildemann, and V. A. Bronnikov, "Individual prediction of motor activity development of patients with a children's cerebral paralysis on the basis of statistical analysis approaches," *Russian Journal of Biomechanics*, vol. 14, no. 2(48), pp. 67–75, 2010.
[17] A. V. Wildemann, A. A. Tashkinov, and V. A. Bronnikov, "Multivariate method of individual motor activity prognostication," *Information Technologies and Computer Systems*, no. 3, pp. 79–85, 2010 (Russian).

[18] S. Simons, "Abstract Kuhn-Tucker theorems," *Journal of Optimization Theory and Applications*, vol. 58, no. 1, pp. 147–152, 1988.

[19] B. Y. Sohn and G. B. Kim, "Detection of outliers in weighted least squares regression," *Journal of Applied Mathematics and Computing*, vol. 4, no. 2, pp. 441–452, 1997.