Future Generation Computer Systems 27 (2011) 139–147

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Web based prediction for diabetes treatment

Lucio Grandinetti *, Ornella Pisacane

Dipartimento di Elettronica Informatica e Sistemistica, Università della Calabria, Via P. Bucci, 41C, Arcavacata di Rende (CS), Italy

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form
27 July 2010
Accepted 8 August 2010
Available online 14 August 2010

Keywords:
Medical information system
Optimization
Decision support system
Web based services

ABSTRACT

Diabetics need continuous support during their therapy because the clinical treatment could be improved in the medical cures;the doctors need to share the new information and to support their patients in an interactive way. In this paper, we describe a web based software tool able to perform predictions on the future glycaemia level of the patients for a specific time horizon; to perform what-if analysis, showing how the glycaemia level could vary if some parameters are changed (either in the diet of the patient or in the clinical treatment); to share information and suggestions, creating a virtual community between specialists and patients; to inform users about their health condition, by a classification algorithm and to allow the patient to maintain a daily diary and to know about the evolution of the health condition. The system merges software engineering and Operation Research methods and could become a valid interactive support for patients and specialists.

1. Introduction

Diabetes is a pathological health condition occurring when the human body is not able to produce or use properly the insulin hormone, which is responsible for controlling the glucose production by the liver and reaching the tissues absorbing it; the direct consequence is an excessive presence of sugar in the blood. Two different types of diabetes exist: 1-type diabetes constitutes about 10% of diabetic patients and is characterized by an absolute absence of insulin, where usually the patient has to inject it every day throughout her life, and 2-type diabetes (about 90% of diabetic patients) characterized by an inability of the human body to properly use the produced insulin. In this case, the patient could manage the problem with an appropriate diet. According to the actual statistics of the American Diabetes Association, 23.6 million people in the United States (about 8% of the population) are affected by diabetes and from 2005 to 2007 the percentage of patients has increased by 13.5%. Moreover, diabetes care costs are about 27 billion US dollars annually for directly treating diabetes, 58 billion US dollars for taking care of the chronic complications attributed to diabetes, and 31 billion US dollars for general medical costs. Most of these costs are attributed to hospital patient care (50%), diabetes medication and supplies (12%), retail prescriptions for diabetes complications (11%) and physician office visits (9%). Information Technology as a contemporary and widespread technology is exploited in health care as well, so that the patient is able to measure at home by herself the actual glycaemia level, to control it in specific hours during the day (for example after eating) and, thus, to maintain a reliable daily diary of these values to receive feedbacks from the doctor.

The aim of this work is to design a web service application to allow patients a real-time interaction with doctors, to share information with other patients and maintain an electronic daily diary in order to reduce general medical expenses. For example, the system is able to give suggestions on a patient's daily diet on the basis of the necessary amount of each nutrient; it allows doctors to support and assist patients during the delicate processes of diagnosis, prognosis and therapy and it informs them on new clinical cases and cures.

The remainder of this paper is organized as follows: in Section 2 a brief literature overview is presented; in Section 3 the data classification problem is considered, describing the relative algorithm; in Section 4 the methods used for performing predictions are described; in Section 5 how the what-if analysis is performed is described; in Section 6 the web application for supporting diabetic patients is described by showing a case study and giving some implementation details and finally Section 7 concludes the paper.

2. Related work

Health Care (HC) represents the treatment and the prevention of pathology and illness. It is one of the largest industries in the world [1,44]. Just to give an idea, HC costs have been rising for several years. Expenditures in the United States on health care surpassed $2.3 trillion in 2008, more than three times the $714 billion spent in 1990, and over eight times the $253 billion

* Corresponding author.
E-mail address: lugran@unical.it (L. Grandinetti).

0167-739X/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2010.08.001

spent in 1980 [2]. Therefore, managers are always interested in funding the most effective HC system using a limited number of resources and minimizing costs [42]. Operations Research (OR) could be used to reach these goals at different levels: how to divide HC resources among patients categories, what portion of the total costs are to be paid by government, how to manage the cost of services and prices for drugs and vaccines, etc. (*Strategic planning*); what goods and services have to be provided and how to allocate resources among them (*Tactical planning*); how to design the HC supply chain, how to allocate and schedule medical staff, how to forecast demand, etc. (*Operational planning*). Of course, different challenges exist. For example, we are not sure whether the data used is relevant and valid or if the models defined are a correct representation of goals and constraints. From a general point of view, a clinical decision making process is made up of three steps (*Diagnosis*, *Prognosis* and *Therapy*) based on the identification, measure and interpretation of patient data. Consequently, a *Clinical Decision Support System* (CDSS) can be seen as a computer system able to provide reminders, advice and/or interpretation to a specific user at a specific time. Then a CDSS can be seen as a significant part of the clinical knowledge management technologies [3] supporting the clinical process and use of knowledge for diagnosis and investigation through treatment and long-term care. We summarize the most important functionalities as follows [4]: Administrative, supporting clinical coding and documentation, authorization of procedures and referrals; Decision support, related to clinical diagnosis and treatment processes by promoting the use of best practices; Managing clinical complexity and details, keeping patients on chemotherapy protocols and preventive care; Cost control, monitoring medication orders avoiding unnecessary tests. In general, an appropriate computerized decision system should be designed based on MKBDSS—Medical Knowledge Based Decision Support Systems (see also [5]). This system comprises two main components: a Medical Knowledge Base (MKB) and an Inference Engine. The former represents a set of known clinical experiences accessible electronically by computers. It constitutes a set of relevant data that can be used by users and it must be constantly updated on the basis of new clinical cases and experiences. Usually, doctors update it on the basis of new clinical cases and therapies.

There are a few diabetics risk assessment calculators available online such as medindia [6], which considers input factors such as age, sex, waist girth, daily routine of exercise, family history. Such calculators are very simple and not taking the advantages of historical data of patient [6]. In addition, there are many tools provided by medindia such as diabetes, paediatric, height–weight tools, clinical, cardiac, HIV-AIDS risk calculators, nutrition and pharma tools. Although the tools are interesting to know some possibilities of health status, they are not using any knowledge management techniques or data mining approaches for prediction.

American diabetics association [7] has used virtual reality to fight against diabetes. The study of 1-type diabetes with computer simulator programs named *Silico*, and *Physiolab*. By using these tools, a scientist can study 1-type diabetes without enrolling people into a clinical trial. In addition, with the help of new computing technology, people with 2-type diabetes can determine their chances of having serious diabetes complications, such as a heart attack or stroke, just by spending a little time on their personal computer. Like other fields of science, computation and data are two basic needs of healthcare as well. For instance, searching for new drugs in pharmaceutical companies and new viruses, modeling of disease outbreak, image processing in medical imaging, bioinformatics, biomedical and determining treatments for patients are some simple and complex examples of Grid [38,40,43] and cloud computing [8] applications in the field of healthcare. In addition, doctors are able to access heath data records transparently regardless of data's place. In all, collaboration and resource sharing that is provided by Grid is very helpful for doctors, patients and other healthcare entities.

Atbrox [9] has created a software that helps individuals with dyslexia to improve their written communication. For a dyslexic person, standard word processing spell checkers are often insufficient. In order to build a statistically accurate language model, there was one fundamental rule: the more data you have, the better. To build this tool, Lingit [9] turned to Atbrox, a Norwegian based company focusing on data mining or data analysis and cloud based solutions with map-reduce and Amazon AWS. The Laboratory for Personalized Medicine (LPM) at Harvard medical school [9] uses preventive healthcare for individuals based on their genetic characteristics by creating models and simulations to assess the clinical value of new genetic tests. To overcome the difficulty of finding enough real patient data for modeling, LPM creates patient avatars, i.e., literally "virtual" patients. The lab can create different sets of avatars for different genetic tests and then replicate huge numbers of them based on the characteristics of hospital populations. Here a solution is given with the help of cloud computing technology for an efficient way to manipulate many avatars, sometimes as many as 100 million at a time. In addition to being able to handle enormous amounts of data, this system can be used by the postdoctoral researchers to find the scope of the genetic risk situation, to determine the appropriate simulation and analysis to create the avatars, and then to quickly build web applications to run the simulations, rather than spending their time troubleshooting computing technology.

Pathwork Diagnostics [10], a molecular diagnostics company, uses machine learning algorithms to analyze and to aid oncologists in the diagnosis of hard-to-identify cancer tumors. Unfortunately, the processing of these models is a highly compute-intensive task. Tens of thousands of models must be processed to find the best model to produce the diagnostic report. While the tests are highly parallelizable, the computation can still take weeks or months using a mid-size high performance computing (HPC) resource, such as a 64-node cluster.

Nimbus Health [11], a Seattle based startup, helps doctors and hospitals to save money by enabling healthcare providers to share medical records with patients in an easy, online, and secure fashion. While the healthcare industry has been slow in transitioning from paper to electronic medical records, Nimbus' transition to the cloud has been anything but slow. By using Amazon Web Services (AWS), Nimbus Health has been gaining the benefits of scale, flexibility, and cost.

Wellness checkpoint [12] Health Risk Assessment (HRA), used by many multinational corporations to track and improve the overall health metrics of organizations. The HRA System is based on questionnaires, in context with the business. This service is available from any PDA device, and has access to global benchmark data.

A real example of an e-Health project that is using full Grid services such as compute, data, security and information is MEDICUS (Medical Imaging and Computing for Unified Information Sharing) [13]. Among others, this project provides services such as teleradiology, image archiving, distributed warehousing and drug discovery. Unlike HealthGrid projects that rely on ad hoc services, this project uses Globus Toolkit solutions to provide services with faster deployment options in agile approach.

One of the main differences between HealthGrid and other Grids like Open Science Grid is that the community behind healthcare system consists of everybody within a health domain that, for example, can be all the people in a country. Therefore, millions of Grid users will be available in this domain and this is a challenge for all Grid services especially security services. While Grids offer the field of medicine numerous benefits, however its application to healthcare is challenging and not straightforward. The SHARE project (Supporting and structuring HealthGrid Activities and Research in Europe) [14] produced a roadmap that identified major challenges on the road to the widespread deployment of HealthGrids.

caBIG (cancer Biomedical Informatics Grid) [15] is a US based healthcare project. It works to provide a collaborative information network for cancer research, however components of caBIG are widely applicable beyond cancer as well. Software and resources in this project are controlled in a way to be federated based on the Grid anatomy, so that they are widely distributed and interlinked. Openness in development, access to ensure broad data sharing and collaboration, and source codes are the other main principles of the caBIG project. Their mission is to accelerate the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer, ultimately improving patient outcomes. Similarly, the EU-funded project neuGRID [16] provides neuroscientists with an advanced e-Infrastructure to develop and clinically assess new imaging markers of neurodegenerative diseases. The project assists in the diagnosis of these types of diseases while also delivering a series of generic biomedical utilities which can be reused across the life sciences.

In [17], a collaborative decision support system, *ViroLab*, is described. It supplies researchers with an interactive framework and provides the medical professional with appropriate tools and information. Its main goal is to allow for personalized drug ranking, integrating tools, modules, protocols and interfaces in a virtual space where users can collaboratively perform tasks. It is on trial in six hospitals and epidemiology laboratories across Europe.

In [18], the authors propose a wireless sensor grid architecture for monitoring the health status of different groups of patients in order to make more easy analysis, diagnosis, prognosis and drug delivery. In [19], the authors propose *VIVE* (a grid-enabled Volumetric Image Visualization Environment) as an interactive analysis tool for 3D medical images, facilitating diagnosis, surgical planning, therapy evaluation, and remote 3D examination. In [20], the authors present a Grid solution to the simulation and visualization of the human arterial system.

In [1], the authors firstly develop an alternative technique to visualize the inner mucosal surface of the colonic wall; secondly, they discuss the possibilities of adopting high performance computing and networking to support virtual reality medical applications. In [21], the authors analyze requirements necessary for a telemedical computing infrastructure. They also compare them with requirements found in a typical metacomputing environment. Thus, they show that metacomputing environments can be used to enable a more powerful infrastructure for telemedicine. Finally, in [22], the authors present and describe *EUROMED-ETS*, a project to secure telemedicine applications over the world wide web.

3. The classification problem

One of the most relevant aspects is related to the data classification problem. In this application, classification trees are used to solve this problem. A classification tree has attributes, internal nodes, branches and leaves. In the case of diabetics treatment, attributes consist of patient data (age, weight, blood pressure, etc.); internal nodes represent tests on attributes; branches represent possible results for the test performed on the parent node and, finally, leaves represent a specific class or a probability distribution for that class (e.g., a diabetic user? If yes, what type of diabetes?). The tree is built from a set of examples and it is used for classifying future samples. It was originally developed by Ross Quinlan [23] and is based on the Concept Learning System (CLS) algorithm whose pseudo-code is described in Algorithm 1, assuming to have a set of training instances S. The feature F to be selected is, generally, chosen by the expert.

Algorithm 1 CLS outline

```
if all the instances s ∈ S are positive then
  create YES node
else
  if all the instances s ∈ S are negative then
    create NO node
  else
    select a feature F with values v_1, ..., v_n and create a decision
    node
  end if
end if
```

The ID3 algorithm improves the CLS algorithm by introducing a feature selection heuristic. In particular, it selects the attribute A that best separates a given set. If it perfectly classifies the set then stop; otherwise it recursively operates on the n subsets to get their best attribute. It uses a greedy search: it picks the best attribute and never looks back to reconsider earlier choices. In addition, the ID3 algorithm selects the best attribute using a statistical property called *Gain* and that measures how well an attribute separates the examples into the set. The one with the highest information is selected. Before defining it in more detail, the *Entropy* concept has to be introduced.

In the following let $|S| = c$:

$$\mathrm{Entropy}(S) = \sum(-p(I) * \log_2(p(I))) \tag{1}$$

where $p(I)$ is the proportion of S belonging to the class I and the sum is defined over c. Thus, if S is a collection of 10 patients, for example, 4 male and 6 female:

$$\mathrm{Entropy}(S) = -4/10 * \log_2(4/10) - 6/10 * \log_2(6/10) = 0.971.$$

Entropy(S) will be equal to 0 if all the members of S belong to the same class and thus the data are perfectly classified. In general, $0 \leq S \leq 1$ and $S = 1$ implies totally random data. The *Gain*, instead, is defined on a couple (S, A) where A is an attribute:

$$\mathrm{Gain}(S, A) = \mathrm{Entropy}(S) - \sum((|S_v|/|S|) * \mathrm{Entropy}(S_v)) \tag{2}$$

where the sum is defined on each value v of all the possible values of the attribute A and S_v is a subset of S for which A has value v. For example, if: $S = 14$ patients; the attribute $A =$ age has two values "$\geq$30" and "$<$30"; there are 8 occurrences of Age = "$\geq$30" and 6 of Age = "$<$30"; for "$\geq$30", 6 are male and 2 female and for "$<$30", 3 male and 3 female:

$$\begin{aligned}\mathrm{Gain}(S, \text{“}\geq 30\text{”}) &= \mathrm{Entropy}(S) - 8/14 * \mathrm{Entropy}(S_{\text{“}\geq 30\text{”}}) \\ &\quad - 6/14 * \mathrm{Entropy}(S_{\text{“}<30\text{”}})\end{aligned}$$

where:

$$\mathrm{Entropy}(S_{\text{“}\geq 30\text{”}}) = 0.811 \qquad \mathrm{Entropy}(S_{\text{“}<30\text{”}}) = 1$$

so, it implies:

$$\mathrm{Gain}(S, \text{“}\geq 30\text{”}) = 0.093.$$

For giving an example in diabetic treatment, S could be a set of 100 patients that could be divided into two categories according to their sex: male and female. After that, according to their age; then their weight and height and finally their actual glycaemia level. Thus, the initial node is made up by all the patients and the set of the attributes $A = \{sex, ages, weight, height, glycaemia\}$. An internal node could be "*age* $<$ 30" and a branch could be made up by "40" patients. This means that 40 patients are less than 30 years old. The leaves could be three types: "healthy patient", "1-type diabetic" and "2-type diabetic". Considering this general description, one can easily see that the ID3 algorithm is very simple to be implemented and used. In general, its first goal is to replace the expert who would, otherwise, build a classification tree manually. As industry has shown, the ID3 algorithm has been effective. We tested it on a set of virtual patients due to some aspects related to data privacy law.

4. Prediction methods

One of the most important functionalities, offered by the system, is a complete set of standard and advanced methods for performing predictions [41] in diabetic treatment. For example, in some cases, the user could be interested in forecasting the glycaemia level some days in advance. The library takes into account three different types of trend for the glycaemia levels:

1. *Constant*: the user has almost the same glycaemia level over a specific horizon;
2. *Linear*: the user has an increasing glycaemia level overall a specific horizon;
3. *Periodic*: the user has a trend of glycaemia level that periodically repeats itself.

For each of these cases, the system has a set of specific methods for performing forecasts. In the following let T be the current period (i.e., the actual date in which the user wants to forecast); τ the number of periods in advance in which the forecast has to be performed; d_T the data related to the period T and $p_{T+\tau}$ the forecast related to the period $T + \tau$. In order to give a general idea about the three classes and their solving methods, we assume to apply the forecast for one period in advance. This implies that the parameter τ is supposed to be 1. If the trend in the input is constant or almost constant (first class), the library performs predictions considering three different methods (in general presented and considered in the literature): the *elementary technique*; the *exponential smoothing* and finally the *moving average* technique. If the elementary technique is used, the prediction for the next period is assumed to be equal to the value of the current period T:

$$p_{T+1} = d_T. \tag{3}$$

According to the moving average technique:

$$p_{T+1} = \operatorname*{sum}_{k=0}^{r-1} \frac{d_{T-k}}{r} \tag{4}$$

where r is a parameter and identifies the last r data available in the historical report. According to exponential smoothing:

$$p_{T+1} = \alpha * d_T + (1-\alpha) * p_T \tag{5}$$

where α is a parameter usually set to 0.2 or 0.3.

On the contrary, if the trend in input follows a linear or almost linear trend (second class), the library will use one of the following methods: the *elementary technique*; the *Holt method*; the *linear regression* and finally the *moving average* method. The general computational schema, common to the four mentioned methods, is the following:

$$p_T(\tau) = a_T + b_T * \tau, \quad \tau = 1, 2, \ldots. \tag{6}$$

By using the elementary technique:

$$a_T = d_T \quad \text{and} \quad b_T = d_T - d_{T-1}. \tag{7}$$

According to linear regression:

$$b_T = \frac{-\frac{(r-1)}{2}\sum_{k=0}^{r-1} d_{T-k} + \sum_{k=0}^{r-1} k d_{T-k}}{\frac{r(r-1)^2}{4} - \frac{r(r-1)(2r-1)}{6}} \tag{8}$$

$$a_T = \frac{\sum_{k=0}^{r-1} d_{T-k} + b_T \frac{r(r-1)}{2}}{r} \tag{9}$$

where r represents the last r data in the historical report. According to the moving average method

$$a_T = 2 * \gamma_T - \eta_T \quad \text{and} \quad b_T = \frac{2}{r-1}(\gamma_T - \eta_T) \tag{10}$$

where

$$\gamma_T = \sum_{k=0}^{r-1} \frac{d_{T-k}}{r} \quad \text{and} \quad \eta_T = \sum_{k=0}^{r-1} \frac{\gamma_{T-k}}{r}. \tag{11}$$

Finally, according to the Holt method

$$a_T = \alpha * d_T + (1-\alpha)(a_{T-1} + b_{T-1}) \tag{12}$$

$$b_T = \beta(a_T - a_{T_1}) + (1-\beta) b_{T-1} \tag{13}$$

setting $a_1 = d_1$ and $b_1 = 0$.

About the third class (periodic trend), the library is able to invoke the following method: the *elementary technique*; the *revised exponential smoothing* and finally the *Winters method*. In this last class, an important parameter is represented by M identifying the length of a cycle after which the trend will repeat itself. For example, according to the elementary technique

$$p_T(\tau) = d_{T+\tau-M}, \quad \tau = 1, 2, \ldots, M. \tag{14}$$

For forecasts with length greater than M

$$p_T(kM + \tau) = d_{T+\tau-M}, \quad \tau = 1, 2, \ldots, M; k = 1, 2, \ldots. \tag{15}$$

For the details of all these methods, the revised exponential smoothing and the Winters method, the reader is referred to [24,25].

Because each prediction could contain some errors, one of the most important problems that has to be solved is related to accuracy with the aim of selecting the best method for a specific trend.

The accuracy of a prediction is evaluated by applying the selected method in the past in order to compute the errors that could be done. Starting from these observations, some measures can be detected: the *mean absolute deviation*; the *mean absolute percentage deviation* and finally the *mean squared error*. These three measures can be used for defining a comparison between the different methods used for forecasting.

To each user, a historical report is associated containing all the glycaemia levels recorded in the past. This information will be used by the system for deciding the specific internal class (constant, linear or periodic trend). Thus, the system will apply each of the methods belonging to one of the three mentioned classes. In order to select the best policy for the specific trend, the system will compute one of the previous measures (in general the mean squared error) for each method. After choosing the best policy, the system will plot the trend of the forecast in order to allow the user to see graphically what happens in the future on her glycaemia level.

Forecasting the glycaemia level can also help the user to control the diet and to preserve the body from eventual pick of this value. This allows the user to constantly control the glucose rate in the blood and to eventually modify diet and medical cures. All the technical aspects are treated directly by the system and the user can just visualize the results. The main steps performed by the system for forecasting the glycaemia level of a specific user can be summarized as in Algorithm 2 using a pseudo-code.

5. What-if analysis

The most innovative part of our application is to allow the users to perform the what-if analysis. In fact, both patient and doctor, using a specific simulation model, can know what happens to the glycaemia level if some clinical parameters change. For example, a patient could be interested in knowing how much the glycaemia level can vary if something is modified in the diet or after an insulin dose; a doctor could be interested in knowing if a patient can have some benefits by varying the glucose input during a day (i.e., increasing or decreasing the quantity of sugar involved in the diet).

Algorithm 2 Selecting the best prediction method

1: Read all the hystorical data of the user;
2: Determine the specific trend followed by the data;
3: Apply all the methods for this specific trend;
4: Select the best method using the Mean Square Error technique;
5: Plot the results

In a general approach, to model how the glycaemia level reacts to an insulin dose, the human body is divided into some compartments. Each of them defines the dynamic behavior of the liver, the kidneys, the pancreas, the organs and the nervous system.

The body could be divided into three zones: the blood capillaries, the tissues and the interstitial fluid.

Mathematically, this compartment model consists of ordinary nonlinear differential equations. In the following, x_1, x_2 and x_3 represent the glucose concentration in the blood (mg/dl), the absorption activity of glucose (min^{-1}) and the insulin concentration in the blood (μU/ml), respectively.

Moreover x_{1B} and x_{3B} are respectively the basal value of glucose (mg/dl) and the basal value of insulin (μU/ml) while $d(t)$ and $u(t)$ are respectively the external source of glucose (min^{-1}(mg/dl)) and the insulin dose (min^{-1}(μU/ml)). Finally, p_1, p_2, p_3 and n are respectively the insulin independent rate of glucose absorption by the tissue (min^{-1}), the decrement rate of glucose in the tissue (min^{-1}), the increment rate of glucose absorption by the tissue per insulin unit above the basal value x_{3B} ($\text{min}^{-2}(\mu\text{U/ml})^{-1}$) and the decrement rate of insulin in the blood (min^{-1}). The mathematical model is described by Eqs. (16)–(19). Eq. (16) models the glucose variation in the blood which is determined by the quantity of glucose given by food or by exercise ($d(t)$) decreased by the insulin independent and dependent variation of glucose (with respect to the basal value). Eq. (17) models the absorption variation of glucose (caused by insulin) in the tissues depends on the difference among the absorption activity of glucose caused by insulin and the glucose decrement in the tissues. Eq. (18) models the insulin variation that depends on the difference of the injected insulin and the quantity that decreases by the factor n. Finally, in Eq. (19) the output variable y represents a measure of glucose in the blood of the patient.

$$\frac{\partial x_1}{\partial t} = -p_1(x_1 - x_{1B}) - x_2 x_1 + d(t) \tag{16}$$

$$\frac{\partial x_2}{\partial t} = -p_2 x_2 + p_3(x_3 - x_{3B}) \tag{17}$$

$$\frac{\partial x_3}{\partial t} = -n(x_3 - x_{3B}) + u(t) \tag{18}$$

$$y = x_1. \tag{19}$$

Our compartmental approach (as also in [26–32]) was validated, analyzed and simulated using the *MATLAB/SIMULINK* environment, considering some specifications taken from real life scenarios.

6. A web service application for supporting diabetic patients

In real life, patients and doctors have to share data even when they are located all over the world. So, we need to develop a web application [34–37,39] for user interaction. Figs. 1–3 show the general system organization and the actions that patients and doctors can perform.

It also has to be very simple and intuitive for user's lacking both software and computer knowledge. We design a web application for supporting diabetic patients. The most important point is that this type of solution could be optimal because it does not need any installation on the client computer. Moreover, it is worth noting that using this web application, users can monitor interactively the glycaemia level in their body; share scientific databases, online journals, news, information, etc.; be informed on new therapies; communicate interactively (by email, forum, etc.) with other users (in different parts of the world). Based on a specific database, the system can support decisions by doctors and patients and for this reason both are indistinctly referred as users of the application (as shown in Fig. 1).

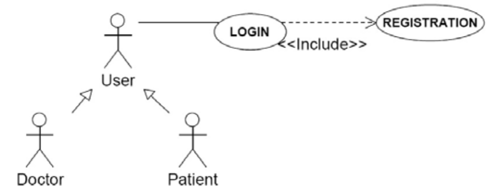

Fig. 1. User.

By implementing the classification algorithm described in Section 3, the system can interpret symptoms and clinical values and thus decide whether or not a patient is diabetic and the type of diabetes. If the patient is diabetic, on the basis of the glycaemia level, the system can support this patient (how to modify the diet, how much insulin is required, what type of analysis the patient has to take, etc.). Moreover it is able to support doctors in their decisions by allowing them to share information with other colleagues, discuss new therapies or support patients.

6.1. System description

In the first phase a user of the system has to perform authentication (see Fig. 1). In modern web applications, in fact, authentication management is one of the first mechanisms for distinguishing between a user and an outsider.

The authentication mechanism is an important element for assuring security: it is not allowed to use simple passwords (i.e., passwords with few characters, typically alphabetical). Generally, in this phase the user sends username and password, as described in the following.

Sending own credentials, the user logs on to the system and the user will be recognized as a diabetic, a new user or a doctor:

- New user: patients register themselves and consequently insert the required clinical values (i.e., their current glycaemia level). In this way the system will be able to classify her as a diabetic (or nondiabetic) patient and (eventually) the specific type;
- Diabetic patient: inserting the current value of glycaemia measured by him/herself, the system will update one's personal electronic diary and decide whether or not the user needs a specific insulin dose or, for example, the patient has to change something in the diet and how.
- Doctor: the doctor can create and use new knowledge (through support diagnosis models, i.e. classification models in order to forecast and establish presence/absence of diabetes for a new clinical case); view detailed data for a specific model (accuracy evaluations, parametric values and so on); eliminate a model; use a model for forecasting; save a diagnosis for a specific clinical case and provides online support for patients during emergency events.

Using the library described in Section 4, both patient and doctor can perform different types of predictions (for a given number of periods, i.e., days or months or years in advance). In particular, the patient will be able to predict the glycaemia level for a given number of periods in advance; the doctor will be able to predict the glycaemia level of a specific patient in the future or the average trend of all the diabetic patients (see Section 4 for details).

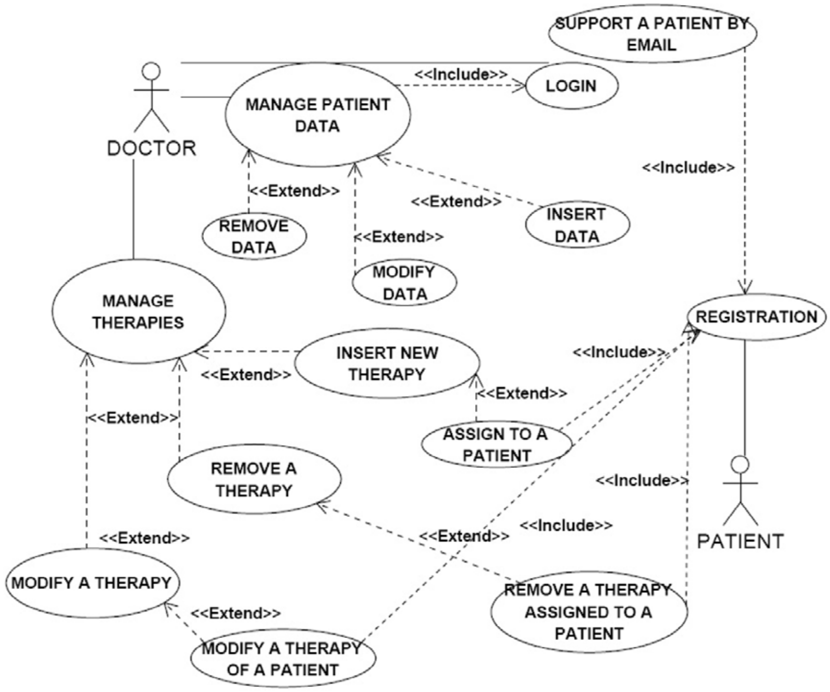

Fig. 2. Doctor.

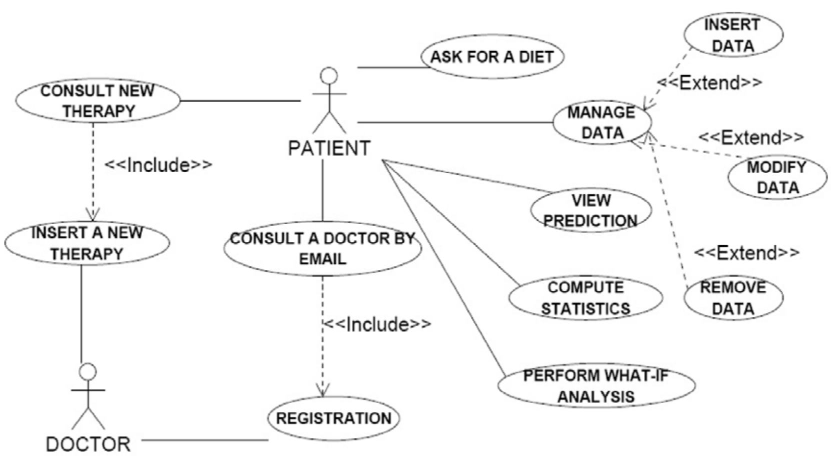

Fig. 3. Patient.

6.2. A case study

In this section we demonstrate our web application functionalities by simple examples to show its most relevant features. It is worth nothing that we cannot describe all the details related to the patient or doctor due to some privacy aspects. In the following, it is assumed that a user is a 2-type diabetic patient and has already performed the login. A menu shows all the possible actions the patient can perform:

- Analysis: new clinical results can be inserted in the database and some comparisons can be performed;
- Clinical consult: a doctor or a specialist already registered in the system can be contacted directly online;
- Diet: inserting the current glycaemia level, a patient can wait for system feedback. As shown in Fig. 4, the user can choose some food and fix their minimum and maximum quantities. For example, the minimum and maximum quantities of chicken are, respectively, fixed by the user (on the basis of own preferences) to 50 and 100 g. In the same table, a red cross allows the user to eliminate a specific food item from the list, while the table below it allows the user to add a new food item. The table on the right, instead, describes the minimum and maximum amounts of nutrients the patient has to receive in a day. These values are fixed by the system on the basis of patient information: glycaemia level, age, weight, diabetes type and so on. After submitting the food preferences, a user can obtain a daily diet which is the result of the optimization of the Diet Problem (DP). The problem is assumed to have $N = \{1 \cdots n\}$ nutrients and $F = \{1 \cdots f\}$ foods; a_{ij} specifies the amount of nutrient i in food j; b_i and d_i are respectively the minimum and the maximum quantities of nutrient i; g_j and h_j are respectively the minimum and the maximum quantities of food j; x_j is the quantity of food j in the daily diet. The objective function minimizes the total amount of sugar in the diet; constraints (20) and (21) impose the minimum and the maximum amounts of each nutrient in the diet, instead (22) and (23) impose the minimum and maximum quantities for each food. In order to solve this problem, an open source java library is used. This is possible since the optimization diet problem is simple: it is solved each day and, thus, the number of variables and constraints is not high and it is a linear problem.

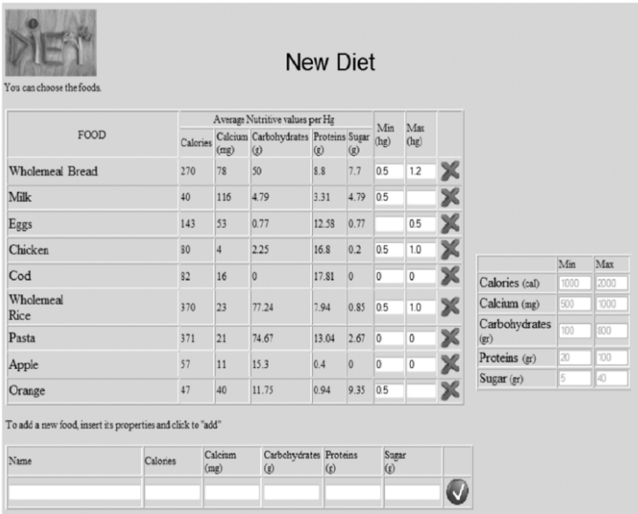

Fig. 4. Diet problem.

DP:

$$\min \quad \sum_{j=1}^{f} a_{\text{sugar},j} x_j$$

$$\text{subject to} \quad \sum_{j=1}^{f} a_{ij} x_j \geq b_i, \quad \forall i \in N \tag{20}$$

$$\sum_{j=1}^{f} a_{ij} x_j \leq d_i, \quad \forall i \in N \tag{21}$$

$$x_j \geq g_j, \quad \forall j \in F \tag{22}$$

$$x_j \leq h_j, \quad \forall j \in F \tag{23}$$

$$x_j \geq 0, \quad \forall j \in F.$$

- Forecast: the patient could be interested in forecasting the glycaemia level trend over a period, specifying it to the system. For example, as shown in Fig. 5, the patient can graphically analyze the forecast of the glycaemia level for 7 days in advance (considering the historical data measured at 8:00 am). In particular, the straight line indicates the reference level of 130 mg/dl and the final dashed part of the curve represents the predictions.
- Insulin level: the patient measures glycaemia in different hours of the day; the user can insert the relative values in order to update her electronic diary and the automatic system could suggest some specific operations to be performed. Moreover, the user can plot the glycaemia trend over a day/month/year and can submit this data to a specific doctor for a periodic control.
- Statistics: the user could be interested in performing some statistics considering personal historical data or general data included in the database. Just to give an idea about this functionality, a doctor could be interested in knowing how many patients could improve their health condition following a specific treatment or how many patients are affected by a specific diabetic type in a specific range of age.
- What-if analysis: the patient could be interested in knowing what happens to the glycaemia level if something changes (i.e., if the patient eats an excessive quantity of sugar in a day). For example, Fig. 6 shows how the user can perform a what-if analysis. First of all, the patient has to provide the system with the current level of insulin and glycaemia; secondly the patient can choose one or more parameters to be varied. If the patient decides to eat more sugar, then the glycaemia level will be varied and consequently so will the insulin dose. The patient can also plot these new trends.

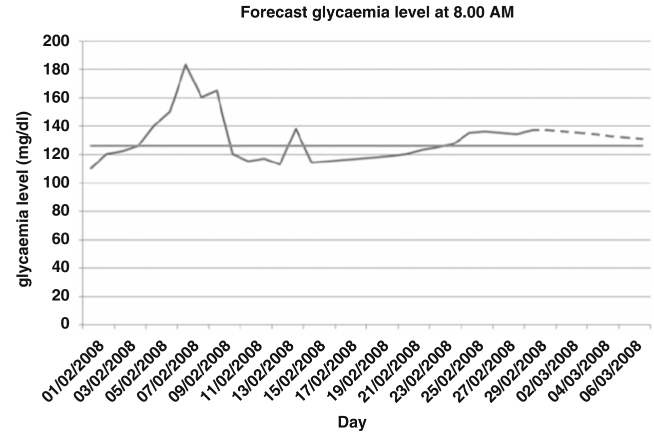

Fig. 5. Predictions at 8:00 am for 7 days in advance.

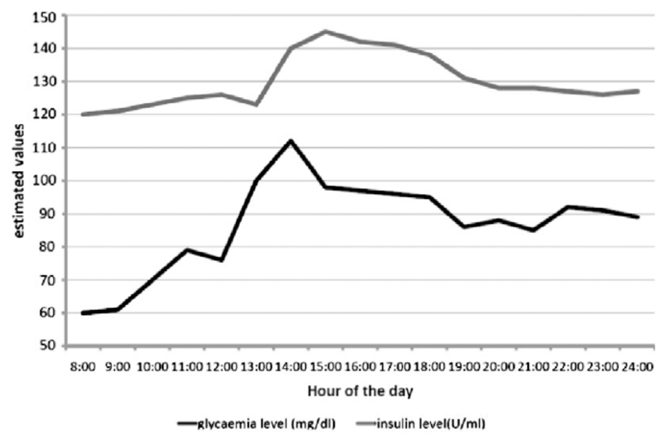

Fig. 6. What-if simulation.

6.3. Implementation

SOA and Data Mining are the underlying technologies used to develop this system. In addition, we adapted this application to be deployed on a WSRF container within the Globus Toolkit framework [33]. WSRF is an implementation of Open Grid Services Architecture. The use cases (Figs. 1–3) show that we reach the interoperability and the other interesting features of Service Oriented Computing, e.g. our services can be deployed on any system easily. Therefore, this application can be used in a Grid system as a consumer or a producer by performing slightly small changes. Among the others, a database system implemented to store patients' and doctors' information, responding to queries and managing data. JDBC library is used to access database operations. Also, WEKA library is used for Data Mining operations. Some OR open source libraries are used in order to solve the Optimization models.

Fig. 7 shows all these components and how they communicate with each other in the global architecture of our application. Our application presents information that are well exposed in a simple way; a set of functionalities that are easy to be used, finally its usability features is of importance to users.

7. Conclusions

In this paper, a web based software tool is described. The system allows diabetic patients to be supported during the therapy and the

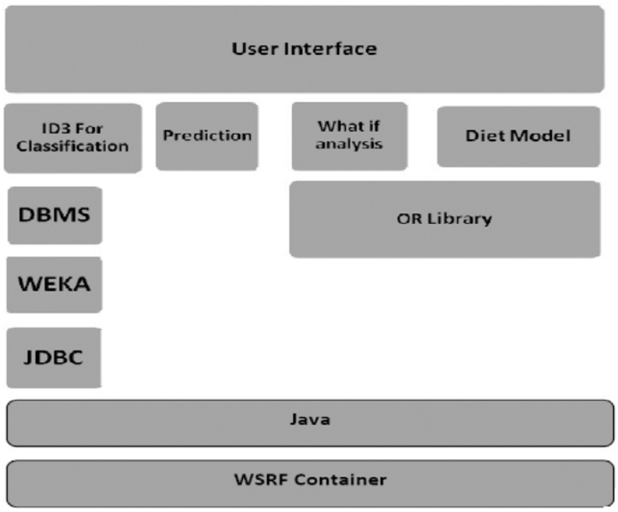

Fig. 7. Software architecture.

clinical treatment; to perform predictions on the future glycaemia level overall a specific time horizon; to perform what-if analysis knowing how the glycaemia levels can vary if some parameters are modified either in the diet or in the cures; to maintain a daily diary and to know about the improvement of the health condition. Dually, the system allows doctors to share information and to know about new therapy and clinical treatment; to interactively support their patients and to perform some useful statistics. Moreover, the system is also able to classify a new user as a healthy, 1-type diabetic or 2-type diabetic patient thanks to the implementation of the ID3 algorithm.

Throughout the application the users can constantly communicate with each other reducing the total medical costs due to diagnosis, prognosis and therapy for the diabetic treatment. It is the result of merging the software engineering techniques with Operation Research methods and it could become a valid interactive support for all its worldwide users.

Acknowledgements

The authors are indebted to Mehdi Sheikhalishahi and Dr. Manoj Devare, for their numerous constructive comments on an early version of the manuscript and for valuable discussions with the authors.

References

[1] A. Marsh, F. Simistirab, R. Robb, VR in medicine: virtual colonoscopy, Future Generation Computer Systems 14 (3–4) (1998) 253–264.
[2] Centers for medicare and medicaid services, Office of Actuary, National Health Statistics Group, National Health Care Expenditures Data, January 2010.
[3] L. Grandinetti, D. Conforti, L. De Luca, CAMD and TeleEEG: software tools for telemedicine applications, in: P. Sloot, M. Bubank, B. Hertzberger (Eds.), High Performance Computing and Networking, in: Lecture Notes in Computer Science, vol. 1401, Springer-Verlag, Berlin, 1998, pp. 64–73.
[4] L. Perreault, J. Metzger, CA pragmatic framework for understanding clinical decision support, Journal of Healthcare Information Management 13 (1999) 5–21.
[5] D. Conforti, R. Guido, Kernel based support vector machine classifiers for early detection of myocardial infarction, Department of Electronics, Informatics and Systems, University of Calabria, Arcavacata di Rende, Optimization Methods and Software, Italy, November 2003 (submitted for publication).
[6] http://www.medindia.net/patients/calculators/diabetes-risk-assessment-calculator-result.asp.
[7] Fighting diabetes with computer technology, Going high-tech: using virtual reality technology to change the future of diabetes. http://www.diabetes.org/news-research/research/recent/hot-topics.html.
[8] I. Foster, Y. Zhao, I. Raicu, S. Lu, Cloud computing and grid computing 360-degree compared, CoRRabs/0901.0131, 2009.
[9] http://aws.amazon.com/solutions/case-studies/atbrox/.
[10] http://pathworkdx.com.
[11] http://www.nimbushealth.com/corporate/Index.html.
[12] http://wellnesscheckpoint.com.
[13] http://dev.globus.org/wiki/Incubator/MEDICUS.
[14] http://initiative.healthgrid.org/the-initiative/share-project.html.
[15] https://cabig.nci.nih.gov/.
[16] http://www.neugrid.eu/.
[17] P.M.A. Sloot, P.V. Coveney, G. Ertaylan, V. Müller, C.A.B. Boucher, M.T. Bubak, HIV decision support: from molecule to man, Philosophical Transactions of the Royal Society, Series A 367 (2009) 2691–2703.
[18] M.P. Rajasekarana, S. Radhakrishnana, P. Subbarajb, Sensor grid applications in patient monitoring, Future Generation Computer Systems 26 (4) (2010) 569–575.
[19] B. Marović, Z. Jovanović, Web-based grid-enabled interaction with 3D medical data, Future Generation Computer Systems 22 (4) (2006) 385–392.
[20] S. Dong, J. Insley, N.T. Karonis, M.E. Papka, J. Binns, G. Karniadakis, Simulating and visualizing the human arterial system on the teragrid, Future Generation Computer Systems 22 (8) (2006) 1011–1017. Also Preprint ANL/MCS-P1344-0506, October 2006.
[21] I. Foster, G. von Laszewski, G.K. Thiruvathukal, B. Toonen, A computational framework for telemedicine, Future Generation Computer Systems 14 (1–2) (1998) 109–123.
[22] D. Polemi, Trusted third party services for health care in Europe, Future Generation Computer Systems 14 (1–2) (1998) 51–59.
[23] J.Ross Quinlan, Induction of decision trees, Machine Learning 1 (1) (1986) 81–106.
[24] G. Ghiani, G. Laporte, R. Musmanno, Introduction to Logistics Systems Planning and Control, Wiley, 2004.
[25] G. Moritz, E. Zeeb, F. Golatowski, D. Timmermann, R. Stoll, Web services to improve interoperability of home healthcare devices, in: 3rd International Conference on Pervasive Computing Technologies for Healthcare 2009, London, UK, Grobbritannien, April 2009.
[26] R.N. Bergman, L.S. Phillips, C. Cobelli, Physiologic evaluation of factors controlling glucose tolerance in man, Journal of Clinical Investigation 68 (1981) 1456–1467.
[27] M. Nomura, M. Shichiri, R. Kawamori, Y. Yamasaki, N. Iwama, H. Abe, A mathematical insulin-secretion model and its validation in isolated rat pancreatic islets perfusion, Computers and Biomedical Research 17 (1984) 570–579.
[28] E. Ackerman, L.C. Gatewood, J.W. Rosevear, G.D. Molnar, Model studies of blood-glucose regulation, Bulletin of Mathematical Biophysics 27 (1965) 21–37.
[29] J.R. Guyton, R.O. Foster, J.S. Soeldner, M.H. Tan, C.B. Kahn, L. Koncz, R.E. Gleason, A model of glucose–insulin homeostasis in man that incorporates the heterogeneous fast pool theory of pancreatic insulin release, Diabetes 27 (1978) 1027–1042.
[30] J.T. Sorensen, A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes, Ph.D. Thesis, Department of Chemical Engineering, MIT, 1985.
[31] E.D. Lehmann, T. Deutsch, A physiological model of glucose–insulin interaction in type 1 diabetes mellitus, Journal of Biomedical Engineering 14 (1992) 235–242.
[32] R.S. Parker, F.J. Doyle III, N.A. Peppas, A model-based algorithm for blood glucose control in type I diabetic patients, IEEE Transactions on Biomedical Engineering 46 (2) (1999) 148–157.
[33] I. Foster, C. Kesselman, The globus project: a status report, Future Generation Computer Systems 15 (5) (1999) 607–621.
[34] A. Ankolenkar, et al., DAML-S: a semantic markup language for web service, in: Proceedings of SWWS'01, Stanford, USA, August 2001.
[35] I. Foster, C. Kesselman, J.M. Nick, S. Tuecke, Grid services for distributed system integration, IEEE Computer 35 (6) (2002).
[36] M. Hori, M. Ohashi, Applying XML web services into health care management, in: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, HICSS'05, vol. 6 (6), 2005, p. 155.1.
[37] IEEE computer: special issue on web service, October 2003.
[38] E. Kimura, M. Ohnishi, T. Norihiko, K. Ishihara, Development of nursing care system with web service, in: AMIA Annual Symposium Proceeding, 2006.
[39] F. Leymann, Web services: distributed applications without limits, database systems for business, Technology and Web BTW 2003, Leipzig, Germany, February 2003.
[40] A. Marsh, The creation of a 21st century telemedical information society, Future Generation Computer Systems 14 (1–2) (1998) 3–22.
[41] D.C. Montgomery, L.A. Johnson, J.S. Gradiner, Forecasting and Time Series Analysis, McGraw Hill, 1990.
[42] G. Neale, C. Vincent, M. Woloshynowych, Exploring the causes of adverse events in NHS hospital practice, Journal of the Royal Society of Medicine 94 (7) (2001) 322–330.
[43] P. Ramnarayan, A. Tomlinson, A. Rao, M. Coren, et al., ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation, Archives of Disease in Childhood 88 (2003) 408–413.
[44] J.C. Wyatt, Clinical Knowledge and Practice in the Information Age: A Handbook for Health Professionals, The Royal Society of Medicine Press, 2001.

Lucio Grandinetti Since 1986, Lucio Grandinetti is a full professor at the Faculty of Engineering, University of Calabria (UNICAL), Italy. At the same University, he currently holds the position of Vice Rector (since 1999) and coordinator of the Ph.D. Degree Program in the area of Management Science and Engineering. At the above mentioned University, he was Director of the Department of Electronics, Informatics, and Systems for ten years, as well as member of the University Administration Council and President of the Management Engineering Degree Course. His scientific background is in Electronic Engineering and Systems Science. He is a graduate of the University of Pisa, Italy and the University of California at Berkeley. He has also been a postdoc fellow at University of Southern California, Los Angeles and Research Fellow at the University of Dundee, Scotland. He has been a member of the IEEE Committee on Parallel Processing, and European Editor of a book series of MIT Press on Advanced Computational Methods and Engineering. Currently he is member of the Editorial Board of three international journals. He is author of many research papers in well-established international journals and Editor or co-Editor of several books on algorithms, software, applications of Parallel Computing, HPC, Grids. Since 1994, he has been part of several Evaluation Panels of European Projects within various ICT and Infosociety Programs; he has also been reviewer of many EU Projects in the above Programs. He has been recipient and scientific leader of many European-Commission-Funded projects since 1993 (e.g. Molecular Dynamics Simulations by MPP Systems, EUROMED, HPC Finance, WADI). He has been recipient and scientific leader of several national projects (CNR progetti finalizzati, Grid.it, PRIN). He is one of the founders of the Consortium SPACI (Southern Partnership for Advanced Computational Infrastructures), a pioneering Grid infrastructure experiment in Italy. Currently he is: Director of the Centre of Excellence on HPC established at the University of Calabria by the Italian government; Co-managing director of a Supercomputing Centre jointly established by the University of Calabria and NEC Corporation; Recipient and scientific leader of some European-Commission-Funded projects (among others, BEINGRID); Project leader of a node of the most important European Grid Infrastructure for e-science named EGEE.

Ornella Pisacane Ornella Pisacane was born in Catanzaro (Italy) on August 11th 1981. In 2004 she obtained the Laurea degree in Computer Engineering at the University of Calabria (UNICAL) defending a thesis whose title was 'Parallel Algorithms of Simulated Annealing for simulation–optimization'. She obtained the maximum score of 110/110 with honours. In 2005 she was an assistant professor for the Optimization Laboratory course at UNICAL. In 2006 and 2007 she was an assistant professor for the Logistics course at the same university. From 2005–2006 she was research fellow at the Departement d'Informatique et de Recherche Operationnelle – Universitè De Montrèal – (Canada), collaborating with Prof. L'Ecuyer's research team to solve some Simulation–Optimization problems. In the same period she attended a master's degree course in ''Simulation Stocastique'' by Prof. L'Ecuyer and a spring school on Vehicle Routing at the same university. Her software skills are mainly related to the use of C, C++, Java, JSP and MPI (or PVM for software implementations on clusters). In 2008, she completed her Ph.D. in Operation Research at UNICAL. From April until September 2009, she collaborated at the CIRRELT – Universitè De Montrèal – (Canada) with Prof. J.F. Cordeau for managing supply chains by using Simulation–Optimization approaches. Her main research areas include Grid Computing and Simulation–Optimization problems. She is the author of scientific publications and technical reports, and the co-author of a book.