A Clinical Support System based on Quality of Life Estimation

Brígida Mónica Faria[1,2], Joaquim Gonçalves[1,3], Luis Paulo Reis[1,4], Álvaro Rocha[1,5]

[1]*LIACC – Lab. Inteligência Artificial e Ciência de Computadores,* [2]*ESTSP/IPP – Esc. Sup. Tecnologia da Saúde do Porto/Instituto Politécnico do Porto,* [3]*EST/IPCA – Esc. Superior de Tecnologia/Instituto Politécnico do Cávado e do Ave,* [4]*EEUM – Escola de Engenharia da Universidade do Minho, Departamento de Sistemas de Informação,* [5]*Universidade de Coimbra, Departamento de Engenharia Informática*

Emails: btf@estsp.ipp.pt, jgoncalves@ipca.pt, lpreis@dsi.uminho.pt, amrocha@dei.uc.pt

Abstract - Quality of life is a concept influenced by social, economic, psychological, spiritual or medical state factors. More specifically, the perceived quality of an individual's daily life is an assessment of their well-being or lack of it. In this context, information technologies may help on the management of services for healthcare of chronic patients such as estimating the patient quality of life and helping the medical staff to take appropriate measures to increase each patient quality of life. This paper describes a Quality of Life estimation system developed using information technologies and the application of data mining algorithms to access the information of clinical data of patients with cancer from Otorhinolaryngology and Head and Neck services of an oncology institution. The system was evaluated with a sample composed of 3013 patients. The results achieved show that there are variables that may be significant predictors for the Quality of Life of the patient: years of smoking (p value 0.049) and size of the tumor (p value < 0.001). In order to assign the variables to the classification of the quality of life the best accuracy was obtained by applying the John Platt's sequential minimal optimization algorithm for training a support vector classifier. In conclusion data mining techniques allow having access to patients additional information helping the physicians to be able to know the quality of life and produce a well-informed clinical decision.

***Keywords** – quality of life; cancer; information technologies; clinical support system; data mining.*

1. Introduction

Quality of life (QoL) is the perceived quality of an individual's daily life, that is, an assessment of their well-being or lack thereof. It includes all emotional, social, and physical aspects of the individual's life. In health care, health-related quality of life is an assessment of how the individual's well-being may be affected over time by a disease, disability, or disorder. Quality of Life (QoL) is an important concept to all the individuals and even more for the patients with chronic diseases. In fact, the term QoL has been used more often in social research, as well as the growing use of the term QoL in clinical intervention experiences, particularly in the field of Oncology, Rheumatology and Psychiatry.

Early versions of healthcare quality of life measure simple assessments of physical abilities (for example, the patient is able to get up, eat and drink, and take care of personal hygiene without any help from others) and single/simple measurements (for example, the angle to which a limb could be flexed). However, the current concept of health-related quality of life acknowledges that subjects put their actual situation in relation to their personal expectation, which can vary over time, and react to external influences such as length and severity of illness, family support, among other factors. Thus, health-related quality of life is now usually assessed using multidimensional patient questionnaires covering physical, social, emotional, cognitive, work- or role-related and even spiritual aspects and also a wide variety of disease related symptoms, therapy induced side effects and other factors such as the financial impact of medical conditions.

Another important area in this context is Information Technology (IT) in health, which is fundamental to the success of care management [1], and there are several good examples of IT systems applications especially in Clinical Decision Support Systems (CDSS). However, the significant growth in the use of IT in healthcare is not accompanied by the same proportion in health benefits. The main reason for this is related with the fact that Knowledge Management Systems are not properly integrated. The use of IT enabled a significant increase in the amount of information available, constituting the main reason for the development of the CDSS. The volume and complex nature of the data justifies the use of robust models capable of transforming information into knowledge. The expectation is that by providing relevant data and knowledge in health care, the CDSS's close the gap between evidence and clinical practice [2]. Nevertheless, the daily use of IT

2

with data mining models capable of helping efficiently the clinic decision is not truly robust. Furthermore, the fact that throughout this process the patient was not involved, that is, these systems are fed with clinical and demographic data related to health, but not with the quality of life of patients.

This paper is organized with seven sections; the first one is composed with this introduction. Second section refers the state of art about information technologies in health. The concept of Quality of Life is presented more completely in section 3. Section 4 presents the developed Quality of Life Platform, the objectives, architecture and mathematical model application to access the individuals' QoL scores. Section 5 explains the methodology and section 6 presents the results and their discussion. Finally, conclusions and future work are presented in the last section of this paper.

2. Information Technologies in Health

Some of the scientific developments related with IT were implemented with success, however with very specific target. For instance METABO [3] is a data mining system for predicting the glucose in diabetic patients and ASPREN [4] is a surveillance system related with the Influenza virus. The United States Institute of Medicine has recognized the existence of problems with the quality of health care and for more than a decade suggests the use of IT in health care, including CDSS to improve the quality of care [5]. Since 2004 the adoption of IT in healthcare has grown slowly. Such IT has been a mean to improve the quality of health care, however the interaction and participation of all partners in health care, including patients, is critical to the successful implementation of IT.

A CDSS is a program designed to help health professionals in decision making. These systems have been the subject of intense study over the past decades in the field of health informatics. The progressive increase in the amount of data, information and knowledge to medical practice has been the main reason for the development of these systems. The expectation is that by providing relevant data and knowledge in the health care, the CDSS will close the gap between evidence and clinical practice.

The Electronic Medical Record (EMR) is growing in Europe and United States. Since 2004 the U.S. federal government promotes the importance of EMR, warning however that although the EMR can improve the accessibility and

readability of information, it is unlikely that there are significant improvements in the quality of services without proper use of the Knowledge Management (KM) system containing the CDSS [5]. Some suggested areas of intervention are: Prevention, for example in immunization; Diagnosis, when there is need to compare patients with signs and similar symptoms; Treatment used in drug interactions alert.

The main issue in KM is to manage and evaluate the knowledge establishing strategies that allow the organization to be aware of their knowledge, since they are often not aware of their knowledge, so going to get knowledge in time and at the right time allowing for greater capacity response, either on products or services.

The KM aims to share, store, create and apply existing knowledge to organizational processes, thus representing a greater capacity for organizational response to fluctuations of its internal and external environment, providing a mechanism favoring the relationships between the different partners of the organizations and at the same time promoting responsiveness and adaptation to the environment. However, this is a relatively new concept in Health, and despite its acceptance, its use has been introduced slowly [6]. Since the 90's a pressure on the health sector with regard to accountability, transparency, equal access to health care and optimization of the efficiency/effectiveness related to services has been made. In this context, various world political entities have promoted the development of Quality Management Systems and external evaluations.

The Integrating the Healthcare Enterprise (IHE) is an initiative for the industry and practitioners of Health to promote information sharing on health through computer systems. It also promotes the distribution of specific clinical needs in support of optimal health care, facilitating communication, promoting the correct use of information, with the advantage of being easy to implement. Examples of this intervention is the "Medical Care in General Practice" created in Germany and published by [7] that demonstrated the applicability and technological feasibility of assessing Quality of Life (QoL) routinely in medical practice and a model for education leaders in hospital management.

In France a public utility, established in 1993, called "Groupement d'intérêt Public Carte de Professionnel de Santé (GIP-CPS)" established an Information System Oriented to the Health Care. This organization was born from a joint initiative of

different entities involved in health care in that country with the objective to electronically and securely share medical data: the Government, the health professionals and insures health organizations in order [8].

The Health Intervention and Technology Assessment Program (HITAP) was implemented in Thailand in 2007 as a subsidiary of a semi-autonomous research unit of the Ministry of Public Health [9].

An information system called "Computer-Based Information Systems" (CBIS) was implemented in the Health area in Bosnia and Herzegovina [10].

Also Scotland and Croatia [11] published IT supported models available for Health. Adani et al. [12] presented the implementation of the model of ADAM developed in Israel, associated with Forensic Medicine and specifically tailored for intervention in disasters. Treatment of data and information relating to Pathology were developed in Information System presented by Daniel in 2009 [13].

Deshpande and Ganz, published in 2009 an article, describing the DiNAR: Health Monitoring of IT Systems in Emergency Response, making it more robust to the use of Information Systems (IS) and Information Technologies in supporting medical emergency [14]. Still supporting the medical emergency was developed in Italy, one specific computer model of epidemiological and clinical assessment and monitoring of acute poisoning in the departments of medical emergency that involves the definition of a computer program for toxicological patients, with easy access to the general system of data collection.

In Canada (for Pathology Cardiovascular) benefits of Information Systems are also documented [15]. In 2009 an electronic clinical support system to assist health professionals in managing the overall risk of cardiovascular disease in Australia decision was published. Australia, also marks the implementation of Flu tracking [4] and "National Notifiable Diseases Surveillance System" as surveillance systems able to provide additional information on influenza in the Australian population.

3. Quality of Life

The perception that an individual has about their place in life, dependent on their culture and values is defined as the individual Quality of Life. In the context of health the classification is Health Related Quality of Life (HRQoL) [16].

Currently the indicators of HRQoL are used in health management strategies for managers, economists and political analysts and pharmaceutical companies using the World Health Organization (WHO) measures of HRQoL in some of its departments. Nowadays, HRQoL is an objective in Medicine, used in epidemiological studies, clinical trials, medical practice in health economic studies, planning of measures and strategies and comparisons between them [16].

Therefore it becomes crucial to add to the clinical and socio demographic information the self-assessment of information's patient about their on QoL.

In addition, regular measurement of QoL allows an effective calculation of QALYs (Quality Adjusted Life Years) enabling also the economic evaluation. The QALYs can be used in three different areas:

- Analysis of clinical decision - the QALYs can be used as a numerical measure of the consequences of the last branches of the decision tree;

- Economic evaluation - allows the choice between different alternatives: we choose that alternative which allows a greater gain in QALYs per year;

- Population studies - groups can be compared in terms of life adjusted for quality.

There are many academic studies on this subject but still no routine implementation. For variability of clinical data adds up to a number of factors whose measurement is subjective:

- Tolerance for pain;

- Emotional state;

- Functional performance or social skills.

These are examples of factors that increase exponentially the complexity of the information system. Consequently, robust techniques are required to extract the knowledge in a way that may constitute, indeed, a clinical decision support, while optimizing the health system from an economic perspective and cost management through [17][18]:

- Rational use of drugs/technologies;

- Control of rate and duration of hospital admissions;

- Hospital reduced costs;

- Prevention programs focusing on at-risk groups;

- Evidence-based Medicine;

- Market research;
- Epidemiological knowledge;
- Pharmacoeconomics knowledge;
- Satisfaction of the patient.

In the study, culminating in QoLis (Quality of Life Information System) project implemented in the head unit and Neck IPO-Porto resulted, half a year, in a database with about 3000 patients. The study, conducted in close collaboration with doctors and nurses revealed that important for the characterization of those patients. QoL assessment reports were produced, with alerts for situations considered abnormal in function of patient data and compared with patients of the same group.

4. Quality of Life Platform

The creation of an information system sustained by mining techniques for data standardization, comparison and discrimination for systematic data analysis is a key step to the successful implementation of health IT. This was precisely the assumption in the creation of QoLis platform [19].

4.1. Objectives

The aim of the QoLis platform, currently being tested in an Oncological Institution, is to offer to healthcare professionals a system of knowledge management. This application will integrate clinical information, information on health status and information on the quality of life of the patient. This platform already experimentally implemented allows the fulfilment of several objectives:

- Provide healthcare professionals a set of tools to measure systematically (at least in every appointment) each patient quality of life without influencing the normal functioning of the health unit;
- Convert tacit knowledge (perception of the patient on their quality of life) into explicit knowledge, producing reports with indications on the health problems of the patient;
- Integrate quality of life information, generated from the platform with existing clinical information (if the unit does not have electronic clinical record this can be done on the platform).

In order to have all the elements available in the application the data mining algorithms and statistical models should allow performing:

- • Evaluation of quality of life;
- • Evaluation of health gains;
- • Evaluation of Functional Health Status;
- • Impact assessment of the consequences of the disease;
- • Categorization of patients through data mining process;
- • Comparison of the patients between categories and within each category analyzing deviant values;
- • Determination of the relevant variation in health status, survival factors and quality of life of patients;
- • Prediction of the health, survival and quality of life in terms of certain factors.

Basically the platform allow having a set of application modules distributed over different integrated systems with a more rigorous and supporting procedure to determine the quality of life of an individual.

In conclusion, the objectives consist in better users' understanding, their patterns of behavior, motivations and reactions to therapies prescribed by doctors. A different perspective is the identification and development of mathematical models to explain and characterize the impact of environment variables for the patients' quality of life.

4.2. Architecture of the Platform

The segment of Public Health also aims to increasingly provide quality and efficient services to the population and, therefore, need to follow the technological trends of extracting knowledge generated by transactional information systems or commonly called On-Line Transaction Processing (OLTP). These systems typically have databases on different technologies, platforms and completely isolated. Therefore, there should be a consolidation of these data so that a human being can have information for making a decision. This transformation is to provide useful and reliable information, in an easy and quick access, forming the knowledge and support the process of decision making for offering higher quality services, thus acquiring greater satisfaction and confidence of users of the Health System.

Fig. 1 shows the architecture of the platform takes these principals into account and involves several modules responsible for Quality of Life evaluation.
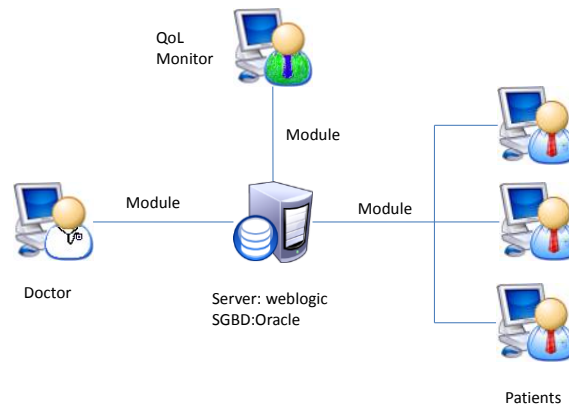

Figure 1. Platform global architecture

This involves the creation of application modules to perform data processing for a set of tools and techniques through the use of data mining learning algorithms and statistics which are able to explore, extract and help highlight patterns in these data with the resulting knowledge discovery.

The application design was intended to be light and efficient in the response (Fig. 2). The mobile modules, developed in ADF Mobile communicate with the frontend which was developed in Oracle ADF in order to communicate with the database. The interface layer with the client was developed in HTML5. Java and JasperReports was required to develop the clinical and statistical reports. To the server page we were used Weblogic on all pages related to the clinical management and Glassfish in the reports that it are a stand-alone modules and which can be used in combination with other applications. The database used was Oracle 12C and the application can run on Linux and derivatives or Windows environment.

The entire presentation layer is rendered on the client, reducing the need for server processing. The technology used was JSP, allowing portability to multiple operating systems. The webpages were developed in JavaScript. The applets are responsible for communication with the client and the CSS contains the elements of graphic style. The business layer contains the rules for access to the server. The software was developed in Java and the servlet's are the components that generate the data to the presentation layer. The language used in SQL data manipulation was embedded in Java code written in the business layer.

The Quality of Life Monitor receives and records clinical data on the platform as the patients respond, also on the platform, to the questionnaires. The report on the patient's QoL is printed at the end of the evaluation process and attached to that is sent to the physician at the time of consultation.

Two questionnaires (Fig. 3) for the evaluation of the quality of life were used: QLQ-C30 e QLQ-H&N35. These questionnaires were developed by EORTC[1], widely used in the evaluation of Quality of Life Related to Health (QoLRH) in oncology patients. The C30 questionnaire is the basis for any oncologic patient, is composed by 30 questions divided in five dimensions, and the H&N35 is the questionnaire used for the specific area of the head and neck composed by 35 questions. The Rasch model was applied in order to have a score to assess the QoL. More details are described in the next sub-section 4.3.
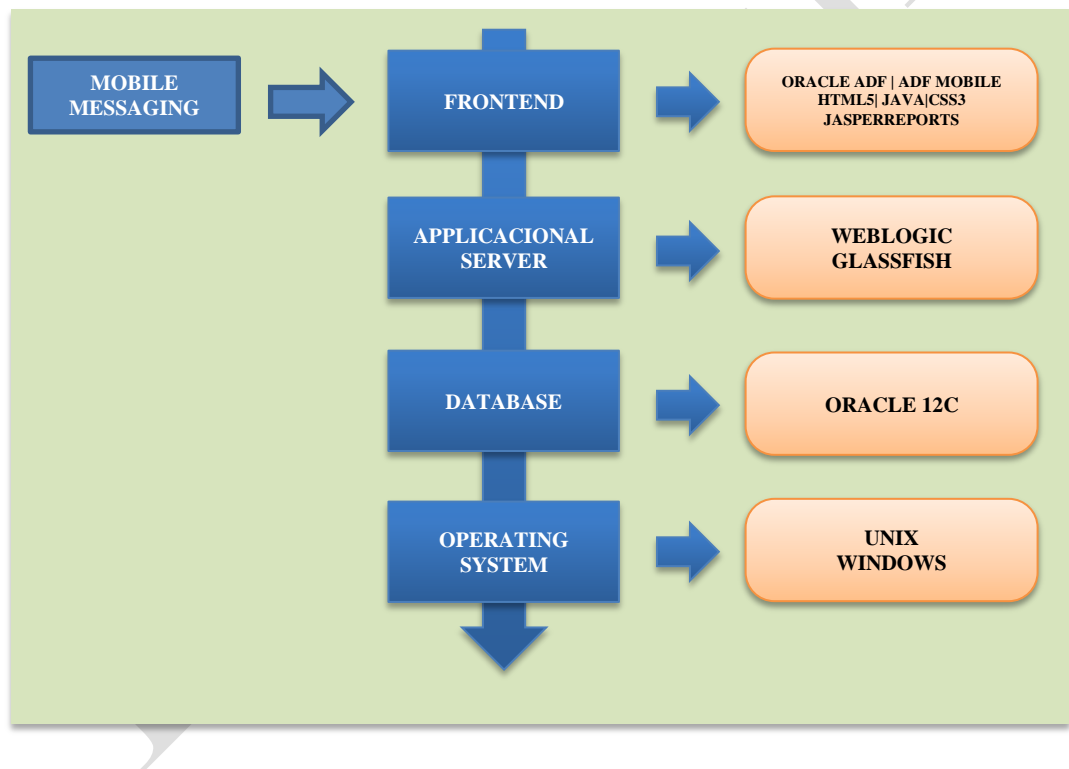
Figure 2. QoLis architecture

It was also tested if there were differences between using the platform and the support of paper in order to assess the results of these questionnaires and there were not statistical differences between the results using these two supports [20].

---

[1] European Organisation for Research and Treatment of Cancer

Figure 3.Quality of Life Questionnaires in the platform

The input data process has been simplified to typify the possible values for each clinical patient variable and respective social demographic data (Fig. 4 and 5). This procedure facilitates the data entry process and reduces typing errors which simplifies the posterior search of information.

Figure 4. Quality of Life Information System – patient data view

Figure 5. Quality of Life Information System – health professional view

The clinical practice of doctors and nurses allowed to determine a set of rules that, depending on the answers given by the patient and the results obtained regarding their QoL, enable the platform to provide guidance on the recommendation for the patient (such as a specific therapy or psychological appointment). The inclusion of these rules in the developed platform allows providing recommendations to the patients in short term in order to improve their quality of life.

4.3. Rasch Model

The Models of Item Response Theory (IRT) is a standard for the design, analysis and scoring of tests, questionnaires and similar instruments measuring abilities, attitudes or other variables [21]. The IRT emerged to supply the needs resulting from limitations of the Classical Theory of Measures (CTM), which was traditionally, and still is, used, in reviews. According to Randall et al. [21] IRT models compared to CTM models have several advantages, namely:

- The statistics are independent of the items of the sample from which they were estimated;
- The results of the examinees are independent of the difficulty of the test;
- The analysis of item accommodates test items consistent with the level of knowledge of the examinees;

- The analysis of the test does not require strictly parallel tests to verify the trust;
- The statistics of the items and the ability of the examinees are both reported on the same scale.

IRT models have potential advantages over the CTM method in the evaluation of responses of self-reported health. These models provide estimates for the items and the latent trait that does not vary with the characteristics of the population. Meaning that measures can be obtained and the variation does not depend on the used instrument or individuals [22].

IRT models typically use a logistic function to link the observed variables to the latent characteristic. In fact, the item parameters can be interpreted as changing the shape of the standard logistic function. Georges Rasch [21], besides this relationship in their model, established a relationship between two other parameters. In fact, the Rasch model is a mathematical expression (1) for the relationship between the probability of success (P) and the difference between the ability of the individual (θ) and the difficulty of the item (β).

$$P = \frac{e^{\theta - \beta}}{1 + e^{\theta - \beta}} \tag{1}$$

From (1) it is possible to obtain the relation in (2):

$$\ln\left(\frac{P}{1-P}\right) = \theta - \beta \tag{2}$$

Therefore, this model assigns a probability of success on the answer to a question in terms of only two parameters to be estimated:
- Efficiency θs of the subject s;
- Difficulty βi of the item i.

In the Rasch model, the scores of the subjects in the test or questionnaire are sufficient to estimate the ability of people. All people have the same score and have the same ability regardless of the responded positively items. Furthermore, item parameters can be estimated without the need to estimate the abilities of individuals [23]. This joint measurement on the scale, the skills of the individuals and item difficulty allows the analysis of interactions between individuals and items. These characteristics fit perfectly the needs of a questionnaire for the assessment of the Quality of Life (QoL), whereas the ability of the subject is his

QoL and that the items are the questions of the questionnaires. In case the answer is not located within the expected value, it is possible to mark it as deviant, comparing the probability of correct response and depending on their QoL (ability) with the response of fact.

5. Methodology

The QoLis platform uses the concepts of knowledge discovery designed to exploit large amounts of data. The objective was to try to replace the questionnaires to measure and quickly indicate the quality of life. This process consists of three basic steps: exploration; model building or pattern definition and validation/verification. In order to extract knowledge from the collected data several steps of knowledge extraction were executed. Next the methods, procedures to collect data and analysis are explained.

5.1. Population, Sample and Data acquisition

The target population is composed by patients with cancer from Otorhinolaryngology and Head and Neck services of an oncologic institution. The sample was composed of 3013 patients.

A space named as *Office of Evaluation of Quality of Life* became the place, in the oncologic institution, where the monitors received the patients, recording clinical data on the platform. At the same time the patients respond, also on the platform, to the quality of life questionnaires. The report on the patients' quality of life is printed at the end of the evaluation process and attached to the file for the physician.

The variables that were acquired and analyzed were indicated by physicians, specialists on cancer from the Otorhinolaryngology and Head and Neck services. Total number of independent variables is 18 and the dependent variable is the total score of the QoL obtained by the questionnaires. Table 1 presents the independent variables.

Table 1. Independent Variables

Independent Variables		
Variables	*Type*	*Options*
Educational Level	Ordinal	0. Analphabetic; 1. 1st cycle; 2. 2nd cycle; 3. 3rd cycle; 4. Frequency of high school; 5. High school; 6. Frequency of BSc; 7. BSc; 8. MSc; 9. PhD

Independent Variables		
Variables	*Type*	*Options*
Marital Status	Nominal	0. Alone; 1. accompanied
Years of Smoking	Ordinal	0. Not a smoker; 1. 1-10; 2. 11-20; 3. 21-30; 4. 31-40; 5. ≥ 41
Number of cigarettes per day	Ordinal	0. None; 1. 1-10; 2. 11-20; 3. 21-40; 4. 41-60; 5. ≥ 61
Years of drinking	Ordinal	0. None; 1. One; 2. 2-5; 3. 6-10; 4. 11-20; 5. ≥ 21
Liters of beer per day	Ordinal	0.None; 1. 1; 2. 2; 3. 3-5; 4. 6-10; 5. ≥ 11
Liters of hard alcohol per day	Ordinal	0.None; 1. 1; 2. 2; 3. 3-5; 4. 6-10; 5. ≥ 11
Size	Ordinal	0. TX; 1. T0; 2. T1; 3. T2; 4. T3; 5. T4
Local metastasis	Ordinal	0. NX; 1. N0; 2. N1; 3. N2; 4. N3
Metastasis distance	Ordinal	0. MX; 1. M0; 2. M1
Histopathological diagnosis	Ordinal	0-9
Tracheostomy	Nominal	0. Yes; 1. No
Type of feed	Nominal	0. Per-os; 1. PEG; 2. SNG
Liters of wine per day	Ordinal	0. None; 1. 1; 2. 2; 3. 3-5; 4. 6-10; 5. ≥ 11
Smoking	Nominal	0. Smoker; 1. Ex-smoker; 2. Non smoker
Years of ex-smoker	Ordinal	0.0-1; 1. 2-5; 2. 6-10; 3. 11-15; 4. ≥ 16
Moment	Ordinal	0. 1st appointment; 1. group appointment; 2. 3 months; 3. 6 months; 4. 9 months; 5. 1 year; 6. 2 years; 7. 3 years; 8. 4 years; 9. 5 years; 10. > 5
Voice prosthesis	Nominal	0. Yes; 1. No

The dependent variable is the Quality of Life Score obtained by the questionnaires and the Rasch model. The variation of the score ranges from -7 to 7 and a classification can be made:

- [-7, -2[– Bad quality of life;
- [-2, 2] – Normal quality of life;
-]2, 7] – Good quality of life.

This classification was presented by the oncologic institution physicians after analyzing the results of the Rasch model and patients. This classification approach is friendlier for the physicians instead of using the QoL score. This means that the obtained score is automatically converted into a simple message with the degree of the quality of life.

5.2. Data Analysis and Algorithms

Descriptive statistics were used for the first data analysis. It was used measures of central tendency such as median and mean and dispersion measures such as standard deviation, minimum and maximum.

15

The technique of multiple linear regression was used to produce a model for predicting the quality of life of the patients. The assumptions of the model were verified: the normality and homogeneity of errors using graphical analysis and the independence of errors with the Durbin-Watson statistic. Variance Inflation Factor (VIF) was used to diagnose multicollinearity. The significance level was established at 0.05. After this analysis, a data mining analysis was performed comparing different algorithms and testing the features that can be associated to the purposed class (Bad, Normal and Good Quality of Life). Appropriated algorithms were tested using the RapidMiner [24]:

- k-Nearest Neighbor – algorithm that for each value inserted finds the closest value and places it in the respective class;

- Naïve Bayes - presents the probability of the accuracy of an object and then applies the Bayes Theorem to produce the classification. This classifier assumes the presence of a particular feature of a class is unrelated to the presence of any other feature. An advantage is that it requires a small amount of training data to estimate the parameters necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix [25];

- Decision Tree (ID3) - This classifier learns decision trees without pruning using nominal attributes which often can also easily be understood;

- W-SMO - John Platt's sequential minimal optimization algorithm for training a support vector classifier [26] [27]. This implementation replaces all missing values and transforms nominal attributes into binary ones. The multi-class problems are solved using pairwise classification (one vs one). Despite being one of the fastest methods for learning support vector machines, SMO is often slow to converge to a solution. Even more when the data is not linearly separable in the space spanned by nonlinear mapping.

The results of the classifiers were compared using all the independent variables to construct the models and only with the variables that could be significant using the multiple linear regression model. A 10 fold cross validation was used to assess the accuracy of the models. The 10 cross validation initially begins with the

division of the dataset into 10 equal-sized partitions. Then a classifier is applied to construct a model using 9 partitions and test on the remaining partition. This step is repeated 10 times, each time using a different partition as the test set.

6. Results and Discussion

After the initial phase of the prototype installation, six months had elapsed since the beginning of the platform installation and it is possible to present results arising from it. The results were organized with a statistical description of the variables that were advised by doctors and in a second stage a deeper analysis was conducted, in order to achieve conclusions about which variables are significant for the quality of life of the patients. The next step consisted in making a comparison analysis about the algorithms to predict the quality of life.

Therefore, Table 2 shows initially the descriptive statistics of the variables. Most of the variables are classified as ordinals, for that reason the central tendency measure was the median and the dispersion was indicated by the minimum and maximum.

Table 2. Descriptive statistics of the Variables

Variables	*n*	**Mean (std)** **or *Median***	**Min/Max**
Educational Level	3013	*Frequency of high school*	Analphabetic/PhD
Marital Status	2942	--	--
Years of Smoking	1785	*31-40*	Not a smoker/≥ 41
Number of cigarettes per day	1826	*11-20*	None/≥ 61
Years of drinking	1811	*≥ 21*	None/≥ 21
Liters of beer per day	1343	*None*	None/≥ 11
Liters of hard alcohol per day	1268	*None*	None/≥ 11
Size	1679	*T2*	TX/T4
Local metastasis	1675	*N0*	NX/N3
Metastasis distance	1677	*M0*	MX/M1
Histopathological diagnosis	2350	0	0-9
Tracheostomy	2851	--	--
Type of feed	2857	--	--
Liters of wine per day	1715	*1*	None/≥ 11
Smoking	2591	--	--
Years of ex-smoker	1390	*2-5*	0-1/≥ 16
Moment	3013	*2 years*	1st appointment/>5
Voice prosthesis	1535	--	--
Quality of Life Score	3013	1.5 (1.24)	-5.66/5.59

A few outliers were encountered in the data revealing that, in future, some messages of confirmation could be introduced in the system, mainly in the extreme options of response, for example the units of beer, hard alcohol or wine. In fact, and after this analysis the units of alcohol were changed from liters to the standard units of alcohol establish by the World Health Organization.

After the analysis of the variables and in order to apply the technique of multiple linear regression the nominal variables were transformed into dummy variables. Although the value of the Durbin-Watson statistic equals 1.293 which indicates a positive auto-correlation, nevertheless the dimension of the sample is big enough for assume an asymptotic independence. Table 3 shows the values of the multiple linear regression model.

Table 3. Multiple linear regression model

Variables	*B*	P value	VIF
Educational Level	-0.110	0.091	1.238
Marital Status	-0.170	0.274	1.220
Years of Smoking	-0.013	0.049	2.428
Number of cigarettes per day	0.002	0.902	2.131
Years of drinking	-0.038	0.959	1.927
Liters of beer per day	-0.139	0.705	2.217
Liters of hard alcohol per day	-0.338	0.443	2.471
Size	-0.158	< 0.001	2.048
Local metastasis	-0.257	0.172	1.755
Metastasis distance	0.043	0.382	1.420
Histopathological diagnosis	-0.528	0.309	1.781
Tracheostomy	-0.170	0.067	3.363
Type of feed	-0.721	0.063	3.476
Liters of wine per day	0.208	0.057	2.818
Smoking	0.021	0.955	1.344
Years of ex-smoker	-0.042	0.560	1.501
Moment	0.038	0.201	1.257
Voice prosthesis	0.619	0.095	1.725
Constant	3.940	< 0.001	--

The VIF values are all lower than 5, which do not indicate problems in terms of multicollinearity. It is also possible to identify by the model the variables that are significant predictors for the Quality of Life: years of smoking (p value 0.049) and size (p value < 0.001). There are also other variables that can be potential predictors, such as: Educational Level; Tracheostomy; Liters of wine per day and

Voice prosthesis. The model is significant ($F(18, 97) = 3.85$; p value < 0.001) and explain 31% of the variability of the Quality of Life. After this analysis several other algorithms were applied and compared using the accuracy (Table 4).

Table 4. Multiple linear regression model

Algorithm	Accuracy (%)	Accuracy with significant and possible predictors (%)	Parameters
k-Nearest Neighbor	55.92	39.83	k=7; measure types: mixed measures; mixed Euclidean distance
Naïve Bayes	63.06	**63.79**	with Laplace correction
Decision Tree (ID3)	42.51	40.26	Criteria: gain ratio; minimal size for split: 4; minimal leaf size: 2; minimal gain: 0.1
W-SMO	**65.58**	58.65	complexity constant C: 1; tolerance parameter: 0.001; epsilon for round-off error: 10^{-12}; random number seed: 1

A comparison about the variables introduced for the creation of the data mining models was also tested: all the variables subjected by the physicians versus the possible predictors for the quality of life. The achieved results show the W-SMO as the best in terms of accuracy (65.58%) for all the initial set of variables, however the Naïve Bayes also produce a similar level of accuracy (63.06%) and the loss in terms of accuracy is lower in comparison with the set of variables that are possible predictors using the multiple linear regression technique.

7. Conclusions and Future Work

Early versions of healthcare quality of life were based on simple assessments of physical abilities and simple measurements of the patient abilities. However, the current concept of health-related quality of life acknowledges that subjects put their actual situation in relation to their personal expectation which can vary over time, and react to external influences such as length and severity of illness, family support, among other factors. Thus, health-related quality of life is now usually assessed using multidimensional patient questionnaires covering physical, social, emotional, cognitive, work related, disease related symptoms, therapy induced side effects and factors such as the financial impact of patient conditions.

The Quality of Life system developed aims exactly to tackle this new concept of quality of life by actively monitoring and estimating the patient quality of life.

The system implements, monitors and promotes the automatic correction of errors and exceptions that may occur during the execution of the different diagnoses and prescriptions. When there is an exception or error, by a method of inference, the application model should have the ability to infer, based on the last situations, the best solution and thus promote their resolution automatically. Moreover, Quality of Life Information System platform answers several questions and requirements in terms of registration and knowledge discovery.

The system provides to the healthcare professionals a multidimensional analysis tools that allows them to analyze all the surrounding variables and cross them with the typical variables arising from questionnaires used to measure quality of life. The platform also provides tools that enable healthcare professionals to apply predictive modeling and simulation of the behavior of users to treatments based on the change of the variables considered. The perception that a given set of treatments, usually expensive, has a negative effect on the quality of life of patients can be used to change the plans, in particular, keeping the patient in less costly palliative care, with gains for the patient quality of life and lower costs to the institution. Moreover, most of the health professionals gave their opinions about the platform and concluded that it will be important for the future all hospitals could provide this service and the idea of having a system that provides information on previous results is very good.

Future work will be related with providing the healthcare professionals with access to information on any mobile device. Availability of the platform in a browser and mobile devices will allow patients to take the questionnaires at home filling them more frequently and thus empowering the system with increased QoL continuous estimation capabilities. In addition the patients, clinicians and nurses will receive immediate feedback of the results. Based on these results clinicians may promptly take appropriate measures to deal with patient QoL changes such as, for example, advancing or postponing a consultation.

References

[1]   Marchibroda JM (2008) The impact of health information technology on collaborative chronic care management. J Manag Care Pharm, vol. 14 (2 Suppl), pp. 3-11.

[2]   Tenório J, Hummel A, Sdepanian V, Pisa I, Marin HF (2011) Experiências internacionais da aplicação de sistemas de apoio à decisão clínica em gastroenterologia. J Health Inf, Jan-Mar, vol.3(1), pp. 27-31.

[3]   Georga E, Protopappas V, Guillen A, Fico G, Ardigo D, Arredondo MT, Exarchos TP, Polyzos D, Fotiadis DI (2009) Data mining for blood glucose prediction and knowledge discovery in diabetic patients: The METABO diabetes modeling and management system, Eng. in Medicine and Biology Society, EMBC 2009. Annual International Conference of the IEEE , pp. 5633-5636.

[4]   Parrella A, Dalton CB, Pearce R, Litt JCB (2009) ASPREN surveillance system for influenza-like illness: A comparison with FluTracking and the National Notifiable Diseases Surveillance System, AuStRALiAN FAMiLy PhySiCiAN Vol. 38, No. 11, pp. 932-936.

[5]   Berner ES (2009) Clinical decision support system: State of the Art. AHRQ Publication, nº 09.0069 – EF. Rockville, Maryland: Agency for Healhcare Research and Quality.

[6]   Balfour DC, Evans S, Januska J, Lee HY, Lewis SJ, Nolan SR, Noga M, Stemple C, Thapar K (2009) Health information technology - results from a roundtable discussion. Journal of Managed Care Pharmacy, vol. 15 (1 Suppl A), pp. 10-17.

[7]   Rogausch A, Sigle J, Seibert A, Thüring S, Kochen M, Himmel M (2009) Feasibility and acceptance of electronic quality of life assessment in general practice: an implementation study. Health and Quality of Life Outcomes.

[8]   Pagliari C, Donnan P, Morrison J, Ricketts I, Gregor P, Sullivan F (2005) Adoption and perception of electronic clinical communications in Scotland. Inform Prim Care., vol. 13 (2), pp. 97-104.

[9]   Tantivess S, Teerawattananon Y, Mills A (2009) Strengthening cost-effectiveness analysis in Thailand through the establishment of the health intervention and technology assessment program. Pharmacoeconomics, vol. 27 (11), pp. 931-45.

[10]   Sivic S, Gojkovic L, Huseinagic S (2009) Evaluation of an information system model for primary health care. Stud Health Technol Inform., vol. 150, pp. 106-10.

[11]   Stevanović R, Stanić A, Varga S (2005) Information system in primary health care. Acta Med Croatica, vol. 59 (3), pp. 209-12.

[12]   Adini B, Peleg K, Cohen R, Laor D (2010) A national system for disseminating information on victims during mass casualty incidents. Disasters, vol. 34 (2), pp. 542-551.

[13]   Daniel C, García RM, Bourquard K, Henin D, Schrader T, Della Mea V, Gilbertson J, Beckwith BA (2009) Standards to support information systems integration in anatomic pathology. Arch Pathol Lab Med., vol. 133 (11), pp. 1841-9.

[14]   Deshpande K, Ganz A (2009) DiNAR: Health monitoring of IT systems in emergency response. Conference Proceedings IEEE Eng Med Biol Soc., 1, pp. 1699-702.

[15]   McFarlane A, Wielgosz A (2009) Strengthening information systems for heart health in Canada. Can J Cardiol, vol. 25 (11), pp. 631-634.

[16]   Pimentel F (2003) Qualidade de Vida do Doente Oncológico. De autor.

[17]   Brereton N, Bodger K, Kamm MA, Hodgkins P, Yan S, Akehurst R (2010) A cost-effectiveness analysis of MMX mesalazine compared with mesalazine in the treatment of mild-to-moderate ulcerative colitis from a UK perspective. Journal of Medical Economics, vol. 13 (1), pp.148-61.

[18]   Moraes E, Campos GN, Figlie NB, Laranjeira R, Ferraz MB (2006) Introductory concepts of health economics and the social impact of the alcohol misuse. Revista Brasileira de Psiquiatria.

[19]   Faria BM, Gonçalves J, Reis LP, Rocha A. (2015) A Platform for Assessing Cancer Patients' Quality of Life, Advances in Intelligent Systems and Computing 354, Springer, vol 2, pp. 51-61.

[20]   Gonçalves J, Rocha Á (2012) Decision support system for quality of life in head and neck oncology patients, Head & Neck Oncology, vol. 4(3), pp. 1-9.

[21]   Randall E, Schumacker P (2005) Item Response Theory. Applied Measurement Associates.

[22]   Castro S (2008) Teoria de Resposta ao Item: aplicação na avaliação de sintomas depressivos. PhD Thesis Univ. Fed. Rio Grande do Sul.

[23]   Mead R (2008) The Measurement Theory of Georg Rasch. Data Recognition Corporation.

[24]   Rapidminer, Available at: http://rapidminer.com/, Consulted in: April 2015.

[25]   Zhang H (2004) The Optimality of Naive Bayes. Faculty of Computer Science, University of New Brunswick, Frederic-ton, New Brunswick, Canada, American Association for Artificial Intelligence.

[26]   Platt J (1998) Machines using Sequential Minimal Optimization. In B. Schoelkopf and C. Burges and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning.

[27]   Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design. Neural Computation. 13(3):637-649.