Development of a Smart Environment for Diabetes Data Analysis and New Knowledge Mining

Eleni I. Georga, Vasilios C. Protopappas, Christos V. Bellos, Vassiliki T. Potsika and Dimitrios I. Fotiadis
Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering
University of Ioannina
GR 451 10, Ioannina, Greece

Eleni Arvaniti and Dimitrios Makriyiannis
Department of Endocrinology
Hatzikosta General Hospital
GR 454 45, Ioannina, Greece

***Abstract*—Diabetes care requires the control of an extensive set of clinical and non-clinical variables which affect the metabolism of glucose in order to prevent acute complications (i.e. hypoglycemic episodes) and to reduce the risk of long-term ones. In this study, we present a clinical information system which records medical (clinical and laboratory) parameters related to Type 1 and 2 diabetes and, mainly, takes a significant step forward towards the collection of lifestyle data. In addition, the intuitive representation and the intelligent analysis of all these multi-parameter data enable the clinician to interpret the status of each patient and support him indirectly in the development of an effective individualized treatment plan.**

Keywords-Type 1 and 2 diabetes; diabetic complications; clinical information system; mobile health devices; data mining

I. INTRODUCTION

The diabetes disease is characterized by persistent hyperglycemia (i.e. increased blood glucose levels), which leads to chronic micro- and macro-vascular complications. Its management is a complex process requiring the collaboration between patients and physicians in order to deal with multiple clinical (e.g. glycemic control) and lifestyle parameters (e.g. diet, physical activity) [1]. Information systems for tracking, monitoring and managing different aspects of the disease have been widely used [2-6]. However, the lack of communication between clinical and patient-centered ones makes it difficult for physicians to access patient's daily routine. In addition, decision support and data mining functionality is occasionaly provided by existing systems [3, 7, 8].

The proliferation of mobile health devices and cloud services has enabled monitoring of personal health information on a more ubiquitous level [9]. Diabetes Pal 2.0 (Telcare®) and TactioHealth (Tactio Health Group) mobile applications are the first to integrate leading health (e.g. glucometers, blood pressure monitors) and lifestyle trackers (e.g. wireless smart activity and sleep trackers) into their monitoring system. On the physician side, data mining of diagnoses, medications, and laboratory results has the potential to reveal unknown disease correlations and predict the incidence of long-term diabetic complications [10-12]. Nevertheless, the efficient integration of self-monitoring data with clinical data can also give a better understanding of the effect of patient's daily context on clinical health outcomes.

The proposed information system (environment) aims at supporting clinicians in the management of Type 1 and 2 diabetic patients by providing tools for data analysis and visualization, extraction of new medical knowledge and treatment planning. It is an innovative intelligent information system which records and processes not only medical parameters but also self-monitoring data related to diabetes.

II. THE PROPOSED SYSTEM

A. System Description

The proposed system has been designed for healthcare professionals and is intended for use in clinical settings. The system provides the following functionalities:

- collects and displays data recorded by lifestyle tracking devices,
- accepts and displays clinical and laboratory examinations coming from routine clinical visits,
- analyses the collected multi-parametric data, provides clinicians with computational tools to extract new knowledge (associations rules, patterns and clusters) and quantifies the progress of the patient,
- allows diabetologists and other healthcare professionals involved in the management of diabetic patients to access and review data and the extracted knowledge,
- provides tools to the physicians for developing or customizing the treatment plan and creating advices / recommendations.

B. The Basic Components of the System

The system, as is shown in Fig. 1, consists of peripheral lifestyle tracking devices and a central system incorporating web-services, a database, the subsystems for data analysis, knowledge extraction and treatment planning, and the graphical user interface. More specifically, the system consists of the following subsystems and modules:

MOBIHEALTH 2014, November 03-05, Athens, Greece

DOI 10.4108/icst.mobihealth.2014.257326

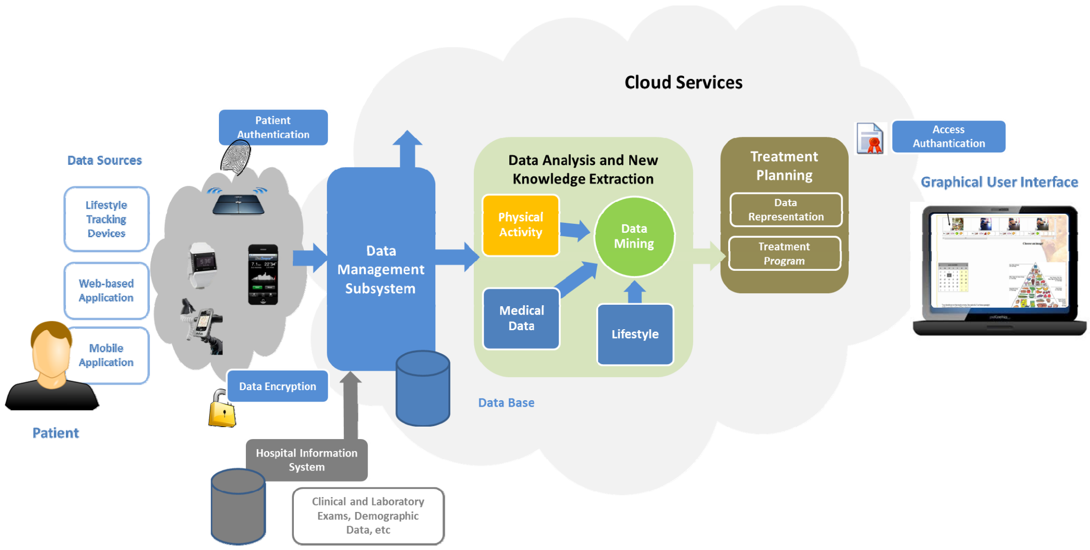

Figure 1. Data flow diagram and system's components.

1) Sources of Self-Monitoring Data: They refer to peripheral devices that patient use in his/her daily life such as smart activity trackers, smart body analyzers, web or mobile applications to record food intake, physical activity or medication, etc. The Withings Pulse, Fitbit Flex and SenseWear Armband activity trackers as well as the Withings Smart Body Analyser will be eventually incorporated into the system. All devices are connected such that measurements are being uploaded to the cloud with minimal user intervention. The output of that module is: parameters describing physical activity (from number of steps, energy expenditure to physical activity type e.g. light, moderate, intense), sleep quality, weight, body composition, calendars and notes.

2) Data Management Subsystem: This subsystem is responsible for collecting data stored on third-party computing clouds as well as on the Hospital Information System (HIS), synchronizing them, storing them to the database located on a dedicated server, and sending them to the other modules of the system.

3) Data Analysis and New Knowledge Extraction: This subsystem is responsible for analyzing the recorded multi-parametric data in order to create the profile of each patient (physical activity quantification, health status, and behavior identification), evaluating the progress of the disease and employing data mining techniques on the basis of this information. These techniques can identify patterns, find relationships describing rules and cluster similar patients with the aim of extracting new knowledge either for an individual patient or for groups of patients.

4) Treatment Planning: It includes tools which allow physicians to develop a treatment plan, establish goals, identify risk factors for a specific patient, customize treatment parameters based on data analysis and data mining and assess whether patient follows the treatment program in his/her daily life.

5) Graphical User Interface: It refers to the user interface through which healthcare professionals can access the above described functions as well as functions for representing data and the results of the analysis in the form of graphs and tables.

C. *Technical Requirements*

The technical requirements of the proposed system are presented in Table I. Regarding interoperability standards, the HL7 [13] is supported at a database level and the ICD-10 [14] is used in the classification of diabetic complications and other diabetes related health problems. To mention that drugs naming is in accordance with the National Drug Organization. The epSOS [15] guidelines are followed in the overall design of patient's medical record. Moreover, JSON lightweight data interchange standard [16] is used for importing data from third parties.

III. Data Analysis and New Knowledge Extraction

The database has so far been populated with 114 Type 2 diabetic patients who are monitored at the Hatzikosta General Hospital of Ioannina from 2 to 28 years (average 10.0±5.1). It consists of 71 women and 43 men whose age ranges from 29 - 85 years (average 68.7±10.4 years). The medical record includes information on glucose control, therapy, clinical and laboratory tests, and medical surveillance of complications or comorbid conditions. In addition, a process of collecting everyday life data is ongoing, in which a group of patients has been equipped with health and lifestyle trackers. Table II presents a high-level grouping of the defined feature set.

An essential component of the proposed system is data visualization, which facilitates the identification of behaviors and trends over time. The visualization tools offered are designed to help the physician to monitor the progress of the disease (e.g. HbA1c), review current or historical examinations and, eventually, learn how patient's habits can impact his health. In particular, international standards (e.g. American Diabetes Association, UK Renal Association) and well-established procedures are followed in the classification and

TABLE I. TECHNICAL REQUIREMETNS OF THE SYSTEM

Database	Microsoft SQL Server 2008 R2
Software Platform	MVC 4.0 (Model – View – Controller)
Programming Language	C#, Html, Javascript
Security Protocols for Communication with Third Parties	HTTPS, oAUTH 2.0
Interoperability Protocols	HL7 (Health Level 7) [13] ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th revision) [14] epSOS (European Patients - Smart open Services) [15] JSON (JavaScript Object Notation) [16]
Hosting Server	Windows Server 2008 R2 Enterprise, SP1

graphical representation of the related parameters. Moreover, simple statistics efficiently represented in tabular or graphical form complement clinical monitoring of diabetes. Fig. 2 shows the simultaneous representation of fasting blood glucose concentration and three hepatic enzymes i.e. AST, ALT and γGT, while Fig. 3 shows physical activity parameters by employing the Withings Pulse device. Both figures have been captured from the current Greek version of the system.

In addition, an integrated set of intelligent tools are developed. Currently, the defined feature space is examined for hidden patterns and similarities by applying unsupervised learning techniques on the dataset of 114 patients [17]. In particular, association analysis i.e. the Apriori algorithm is employed to identify rules relating therapy scheme glucemic control and diabetic complications, while, the k-means algorithm is used to identify groups of patients with similar patterns. Nevertheless, given the inherent sequential nature of diabetes data, techniques able to identify both co-occurrence and temporal dependencies should be studied. The next step is to extend the input of the models by utilizing self-monitoring information on health and lifestyle patterns.

The problem of the estimation of the long-term progression

TABLE II. DESCRIPTION OF THE FEATURE SET

Demographic and Clinical Data	
Patient Profile	Gender, Age, Diabetes duration, Family history, Obesity, Weight, Body Mass Index, Birth weight, Habits (e.g. smoking)
Glucose-Insulin Control	Glycated hemoglobin (HbA1c), Hypoglycemia, Fasting blood glucose, Insulin concentration
Therapy	Oral medication, Insulin therapy
Nephropathy	Serum creatinine and calculated GFR, Microalbumin, Potassium
Retinopathy	Fundus examination, Intraocular pressure
Neuropathy	Skin and foot examinations, Clinical manifestations of autonomic neuropathy
Cardiovascular Disease	Systolic and diastolic blood pressure, Low- and high-density lipoprotein, Triglycerides
Fatty Liver	Abdominal ultrasound, Enzymes: AST, ALT, gGT
Comorbidities	Hypo/hyperthyroidism, Pernicious anemia, Celiac disease
Self-monitoring Data	
Physical Activity	Number of steps, Total distance, Energy expenditure, MET levels, Physical activity levels and duration.
Health and Lifestyle Data	Weight, Body composition, Sleep quality, Diet

of the disease (i.e. HbA1c, incidence of diabetic complications, risk of hypoglycemia) is also addressed in the context of well-established classification techniques of machine learning (e.g. Support Vector Machines, Random Forests) [17, 18]. The applied therapy scheme and patient's overall lifestyle constitute the primary predictors of glucose control; whereas, higher glucose levels are associated with increased risk of complications.

IV. CONCLUSIONS

In this study, we present the functionality of a clinical diabetes information system that enables: (i) the efficient monitoring of clinical, laboratory and self-monitoring data, (ii) the interpretation of patient's status through the intelligent analysis of long-term data, (iii) the extraction of new medical knowledge, and (iv) the personalization of treatment

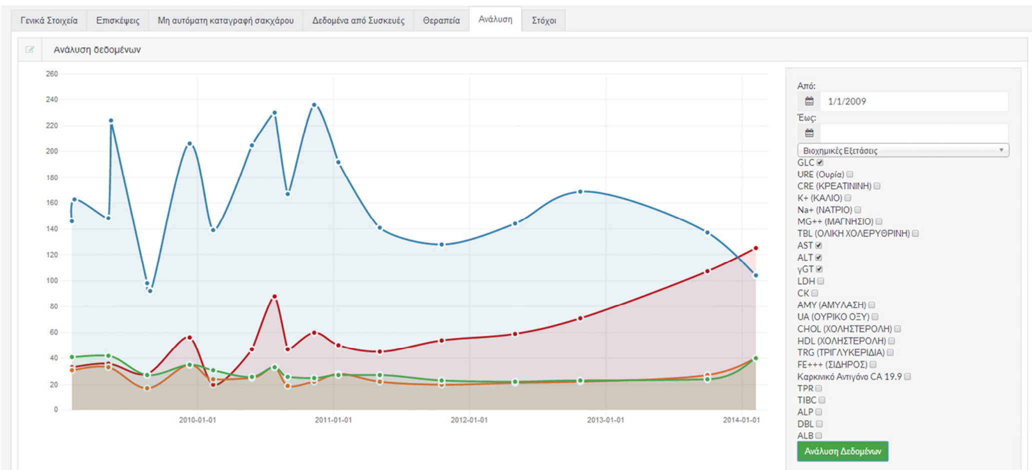

Figure 2. Representation of four biochemical parameters i.e. blood glucose GLC (with blue line) and hepatic enzymes AST, ALT and γGT (with green, orange and red line respectively) for a typical Type 2 diabetic patient.

Ημερομηνία	Βήματα	Απόσταση (m)	Θερμίδες (kcal)	Υψόμετρο (m)	Ήπια Άσκηση (min)	Μέτρια Άσκηση (min)	Έντονη Άσκηση (min)
2014-8-1	8547	8235.02	575.76	25.9	2520	1140	1380
2014-8-4	6625	4994.56	325.62	75.64	3420	1200	240
2014-8-5	11108	10674.85	753.76	20.75	3060	1260	1800
2014-8-6	2982	2550.78	110.63	19.65	1680	360	120
2014-8-7	5823	4771.72	210.95	37.84	3120	780	60
2014-8-8	4381	3653.87	176.79	57.12	1740	1080	180
2014-8-9	5657	4634.428	205.7	31.88	2220	1500	60
2014-8-10	12125	10922.97	455.13	27.18	3240	3840	240
2014-8-11	3089	2551.09	104.93	6.79	2160	60	0
2014-8-12	4236	3540.84	155.37	26.79	2280	240	120
2014-8-13	10006	9389.89	367.68	18.89	1560	3540	180
2014-8-14	12798	12122.76	481.63	24.89	2760	3960	300
2014-8-21	2915	2436.51	111.08	29.24	1620	540	60
2014-8-22	7298	7982.66	440.13	19.07	1620	600	1620
2014-8-23	1138	928.8	38.98	2.34	840	60	0
2014-8-29	1822	1528.87	65.77	9.33	1080	60	60
2014-8-30	3413	2761.24	116.73	4.76	1320	960	0
2014-8-31	1024	856.98	40.93	13.02	660	120	60
2014-9-1	2692	2262.21	103.13	27.42	840	600	60
2014-9-2	3215	2662.45	123.25	30.28	1320	540	120
2014-9-3	3320	2795.88	126.88	29.91	1140	660	120
2014-9-4	6421	5177.47	236.69	42.96	2100	1980	120
2014-9-5	4013	3320.69	151.73	34.35	2640	360	60
2014-9-6	4862	3822.07	167.88	12.65	1680	1560	0
2014-9-7	5783	4842.56	203.71	13.67	2220	1800	60

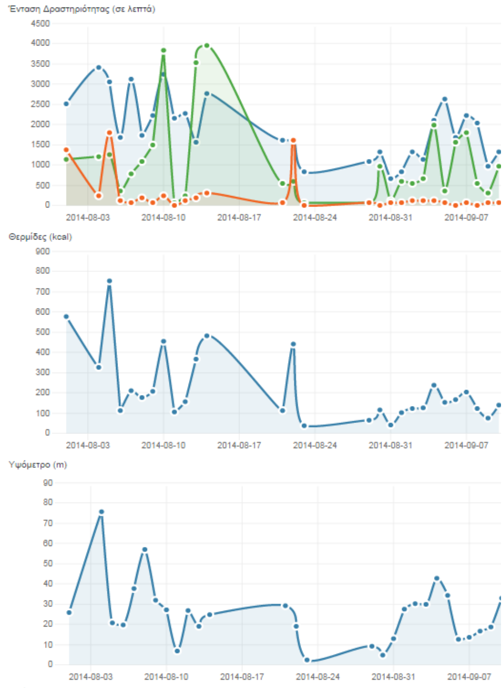

Figure 3. Representation of physical activity parameters e.g. intensity, calories and elevation as acquired from Withings Pulse activity monitor for a typical Type 2 diabetic patient.

recommendations. The extracted knowledge along with existing clinical practice may contribute to the discovery of cause-and-effect relationships in diabetes disease management, resulting indirectly in supporting its clinical and daily care and preventing its complications.

ACKNOWLEDGMENTS

This work is supported by the research project "Development of an Information Environment for Diabetes Data Analysis and New Knowledge Mining" that has been co-financed by the European Union (European Regional Development Fund - ERDF) and Greek national funds through the Operational Program "THESSALY-MAINLAND GREECE AND EPIRUS-2007-2013" of the National Strategic Reference Framework (NSRF 2007-2013).

REFERENCES

[1] American Diabetes Association, "Standards of medical care in diabetes", Diabetes Care, vol. 37, no. Sup 1, pp. S14-S80, Jan. 2014.

[2] R. Jeffery, E. Iserman, and R.B. Haynes, "Can computerized clinical decision support systems improve diabetes management? A systematic review and meta-analysis", Diabet. Med., vol. 30, pp. 739-745, Apr. 2013.

[3] K. Nirantharakumar, Y.F. Chen, T. Marshall, J. Webber, and J.J. Coleman, "Clinical decision support systems in the care of inpatients with diabetes in non-critical care setting: systematic review", Diabet. Med., vol. 29, pp. 698-708, Jun. 2012.

[4] O. El-Gayar, P. Timsina, N. Nawar, and W. Eid. "Mobile applications for diabetes self-management: Status and potential", J. Diabetes Sci. Technol., vol. 7, pp.247-262, Jan. 2013.

[5] P. Klasnja, and W. Pratt, "Healthcare in the pocket: Mapping the space of mobile-phone health interventions" J. Biomed. Inform., vol. 45, pp. 84-198, Feb. 2012.

[6] J.L. Schnipper, C.L. Liang, C.D. Ndumele, and M.L. Pendergrass, "Effects of a computerized order set on the inpatient management of hyperglycemia: a cluster-randomized controlled trial.", Endocr. Pract. vol. 16, pp. 209-218, Mar.-Apr. 2010.

[7] E.I. Georga, et al., "Data mining for blood glucose prediction and knowledge discovery in diabetic patients: The METABO diabetes modeling and management system", Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 5633-5636, 2009.

[8] J.U. Poulsen, et al., "A diabetes management system empowering patients to reach optimised glucose control: From monitor to advisor", Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 5270-5271, 2010.

[9] E.I. Georga, V.C. Protopappas, C.V. Bellos, and D.I. Fotiadis, "Wearable systems and mobile applications for diabetes disease management", J. Health and Technology, pp. 1-12, May 2014.

[10] Y. Huang, P. McCullagh, Black N, and R. Harper, "Feature selection and classification model construction on type 2 diabetic patients' data", Artif. Intell. Med., vol. 41, pp.251-262, Nov. 2007

[11] M. Marinov, A.S. Mosa, I. Yoo, and S.A. Boren, "Data-mining technologies for diabetes: a systematic review", J. Diabetes Sci. Technol., vol. 1, pp. 1549-1556, Nov. 2011.

[12] P.B. Jensen, L.J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care". Nat. Rev. Genet., vol. 13, pp. 395-405, May 2012.

[13] Health Level 7 Standard, http://www.hl7.org/implement/standards/ HL7.

[14] International Statistical Classification of Diseases and Related Health Problems 10th Revision, http://apps.who.int/classifications/icd10/.

[15] European Patients - Smart open Services, http://www.epsos.eu/ epSOS.

[16] The JSON Data Interchange Standard, http://json.org/.

[17] P-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Addison-Wesley, 2005.

[18] C.M. Bishop, "Pattern recognition and machine learning", Springer 2006.