Journal of Diabetes and Its Complications 29 (2015) 691–698

Contents lists available at ScienceDirect

Journal of Diabetes and Its Complications

journal homepage: WWW.JDCJOURNAL.COM

Realization of a service for the long-term risk assessment of diabetes-related complications

Vincenzo Lagani [a,*], Franco Chiarugi [a], Dimitris Manousos [a], Vivek Verma [b], Joanna Fursse [c], Kostas Marias [a], Ioannis Tsamardinos [a,d]

[a] Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, Greece
[b] Department of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, United Kingdom
[c] Chorleywood Health Center, Chorleywood, United Kingdom
[d] Department of Computer Science, University of Crete, Heraklion, Greece

ARTICLE INFO

Article history:
Received 28 November 2014
Received in revised form 6 March 2015
Accepted 17 March 2015
Available online 25 March 2015

Keywords:
Diabetes complications
Clinical decision support systems
Machine learning
Risk assessment models
DCCT / EDIC studies

ABSTRACT

Aim: We present a computerized system for the assessment of the long-term risk of developing diabetes-related complications.
Methods: The core of the system consists of a set of predictive models, developed through a data-mining/machine-learning approach, which are able to evaluate individual patient profiles and provide personalized risk assessments. Missing data is a common issue in (electronic) patient records, thus the models are paired with a module for the intelligent management of missing information.
Results: The system has been deployed and made publicly available as Web service, and it has been fully integrated within the diabetes-management platform developed by the European project REACTION. Preliminary usability tests showed that the clinicians judged the models useful for risk assessment and for communicating the risk to the patient. Furthermore, the system performs as well as the United Kingdom Prospective Diabetes Study (UKPDS) Risk Engine when both systems are tested on an independent cohort of UK diabetes patients.
Conclusions: Our work provides a working example of risk-stratification tool that is (a) specific for diabetes patients, (b) able to handle several different diabetes related complications, (c) performing as well as the widely known UKPDS Risk Engine on an external validation cohort.

1. Introduction

Diabetes is widely recognized as one of the major threats for health in developed and, even more, in developing countries. In particular, diabetes complications represent a relevant burden and source of sufferance for the single patient, and a plague for the whole community in terms of health cost and missed productivity. Stratifying diabetes patients in accordance to their risk of developing complications is usually helpful in order to contrast the negative effects of such diabetes complications. A precise stratification is beneficial in order to (a) design an effective care plan for each single patient and (b) efficiently budget and manage health care centers' economical and human resources.

However, assessing the risk of developing a given diabetes-related complication is not a trivial task. The actual risk is the result of a number of several factors closely interacting, and their interplay is often not fully understood. Health care providers frequently follow large numbers (e.g., thousands) of diabetes patients, and do not have enough human resources (doctors, specialized nurses) to assess the individual risk of each single patient for each possible complication. In order to cope with such difficulties, Clinical Decision Support Systems (CDSS) have recently emerged as software tools for risk assessment and patient outcome prediction in clinical settings (Garg et al., 2005).

In this work a novel computerized tool for the evaluation of the personalized, long-term risk of developing diabetes-related complications is presented. The tool is named LTRA, which stands for long-term risk assessment system. The system accepts a set of clinical parameters describing the health status of a patient (e.g., as extracted from an electronic health record) and computes the patient's probability over time of developing a set of diabetes-related complications. LTRA general architecture is illustrated in Fig. 1. The core of the system is composed by a set of mathematical/statistical predictive models and a module for the management of missing information. The predictive models were derived on the data collected in the Diabetes Control and Complications Trial (DCCT, one of the largest case-control randomized trials to date for

Conflicts of interest: The authors declare that there are no conflicts of interest.
* Corresponding author at: N. Plastira 100, Vassilika Vouton, GR-700 13 Heraklion, Crete, Greece. Tel.: +30 2810 391070; fax: +.30 2810 391428.
E-mail address: vlagani@ics.forth.gr (V. Lagani).

http://dx.doi.org/10.1016/j.jdiacomp.2015.03.011
1056-8727/

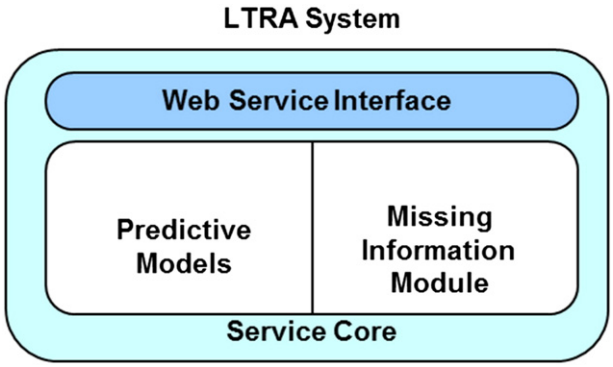

Fig. 1. General architecture of the long-term risk assessment (LTRA) system. The core of the system is composed by a set of predictive models and a module for the management of missing information. The system is deployed with a Web service interface in order to facilitate the interaction with external programs/services.

type I diabetes) (The Diabetes Control and Complications Trial Research Group, 1993) by applying an elaborated machine-learning/data mining analysis protocol (Lagani et al., 2015).

The presence of missing information is a common issue when dealing with clinical records. The missing information module is specifically devised in order to allow the LTRA system to provide reliable risk evaluations even when information required by the predictive models is missing. Whenever a patient's profile lacks one or more clinical parameters, the missing information module reconstructs the multivariate distribution of the missing information, conditional to the values of the available clinical parameters. An average assessment of the risk is then provided by summing over all possible value combinations of the missing clinical parameters.

Finally, The LTRA system has been embedded within a Web service interface, in order to allow remote interoperability and integration in external applications.

LTRA is not the first CDSS to be proposed for the long-term risk assessment of diabetes related complications. A well-known similar application is the UKPDS Risk Engine, where UKPDS stands for United Kingdom Prospective Diabetes Study (Turner & Holman, 1996). The UKPDS Engine calculates the risk over 10 years of experiencing a (fatal) coronary heart disease or stroke on the basis of a set of patient's clinical parameters (age, sex, diabetes duration, glycated hemoglobin, cholesterol profile, smoking status, presence of atrial fibrillation). All the information must be provided in order to obtain an evaluation. Another notable example of CDSS for calculating the risk of developing heart diseases is the Q-Risk system (Hippisley-Cox et al., 2007). With respect to the UKPDS Risk Engine, the Q-Risk predictive model is based on a larger set of clinical parameters, such as family history of heart diseases, treatment information and the presence of comorbidities. Moreover, Q-Risk can assess the risk for both diabetes and non-diabetes patients. Several other predictive models have been developed for assessing diabetes patients' risk of experiencing adverse events, including death (McEwen et al., 2012), macro albuminuria (Lopes-Virella et al., 2013) and retinopathy (LeCaire, Palta, Klein, Klein, & Cruickshanks, 2013). However, no implementation is available for most of these models. Interested readers can find further information in the systematic review presented in Lagani et al..

The LTRA system was developed in the context of the European project REACTION (REmote ACcessibility to diabetes management and Therapy In Operational healthcare Networks). The REACTION project "aims to research and develop an intelligent service platform that can provide professional, remote monitoring and therapy management to diabetes patients in different healthcare regimes across Europe" (http://www.reaction-project.eu/news.php). The LTRA services are integrated into the REACTION clinician portal, which is the interface through which medical doctors can access all the services offered by the platform.

The remaining sections are organized as follows. Section 2 describes in detail the architecture of the LTRA system, while Section 3 illustrates its integration within the REACTION platform. The results of a preliminary usability test and of the comparison with the UKPDS Risk Engine (Stevens, Kothari, Adler, & Stratton, 2001) are reported in Section 4. The Discussion and Conclusions section summarizes and comments the results reported in this work.

2. Architecture of the LTRA system

2.1. The predictive models and their derivation

The predictive models were derived and validated elsewhere (Lagani et al., 2015). The data-driven, machine-learning approach used for building the model is briefly summarized here. A predictive risk-assessment model $f(t,x)$ is a mathematical formula which accepts as input a time-horizon t and a patient profile x, and provides the survival function $S(t|x)$, i.e., the probability of being complication-free until the time t given the specific profile x of the patient. The patient profile $x = \{x_1, x_2, \cdots, x_n\}$ is composed by a set of measurements $x_1, \cdots, x_n$ over n clinical parameters $X_1, \cdots, X_n$ (e.g., smoking status, glycated hemoglobin, etc.). Clearly, the utility of the LTRA component is directly proportional to the predictive capability of the mathematical models that form its core.

2.1.1. The data

The Diabetes Control and Complications Trial database, provided by the National Institute of Diabetes and Digestive and Kidney diseases (NIDDK), was used for deriving the predictive models. The DCCT study run through-out the '90s for about 10 years in the US and Canada; the official documentation of the study states that: "1,441 volunteers, ages 13 to 39, with type 1 diabetes and 29 medical centers in the United States and Canada. Volunteers had to have had diabetes for at least 1 year but no longer than 15 years. They also were required to have no, or only early signs of, diabetic eye disease".[1] Fifty-one clinical parameters collected during the first visit were considered as possible risk factors (see Supplementary Table 1). The information collected during the follow-up visits was used in order to identify the subjects that experienced adverse events (complications). The analyzed data were affected by censorship, i.e., the exact time-to-event for each complication was not available for all the subjects present in the study, since only a subset of the subjects of the DCCT cohort experienced some complications.

2.1.2. Derivation of the models

The following objectives were pursued during the creation of the LTRA models:

1. Inducing accurate models for a set of diabetes-related complications. Diabetes affects multiple organs and systems of the body; consequently, diabetes patients can develop several different complications. Seven different complications were judged highly relevant for this work, either for their incidence or severity: cardiovascular diseases (CVDs), hypoglycemia, ketoacidosis, microalbuminuria, neuropathy, proteinuria and retinopathy.
2. Identifying, for each complication, the minimal-size subset of clinical parameters necessary for optimal prediction. Identifying such subset of risk factors provides intuition into the mechanisms causing the disease, resulting in models easier to verify, understand and visualize. Clinical factors not included in this set are either irrelevant or redundant given the selected ones.
3. Providing an unbiased estimate of the performance of the induced models. Obviously, it is not enough to construct predictive models

[1] http://diabetes.niddk.nih.gov/dm/pubs/control/DCCT-EDIC_508.pdf, retrieved on 27/02/2014.