Expert Systems With Applications 78 (2017) 301–318

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Improvement of newborn screening using a fuzzy inference system

Unai Segundo[a,*], Luis Aldámiz-Echevarría[b], Javier López-Cuadrado[c], David Buenestado[a], Fernando Andrade[b], Tomás A. Pérez[a], Raúl Barrena[a], Eduardo G. Pérez-Yarza[d], Juan M. Pikatza[a]

[a] Department of Computer Languages and Systems, Computer Science Faculty, University of the Basque Country (UPV/EHU), Donostia - San Sebastián, Gipuzkoa, Spain
[b] Division of Metabolism, Cruces University Hospital, Barakaldo, Bizkaia, Spain
[c] Department of Computer Languages and Systems, School of Industrial Technical Engineering, University of the Basque Country (UPV/EHU), Bilbao, Bizkaia, Spain
[d] Division of Pediatrics, Donostia University Hospital, Donostia - San Sebastián, Gipuzkoa, Spain

ARTICLE INFO

Article history:
Received 4 January 2016
Revised 10 February 2017
Accepted 11 February 2017
Available online 15 February 2017

Keywords:
Decision support system
Fuzzy inference system
Inborn error of metabolism
Newborn screening

ABSTRACT

This paper presents a decision support system (DSS) called DSScreening to rapidly detect inborn errors of metabolism (IEMs) in newborn screening (NS). The system has been created using the Aide-DS framework, which uses techniques imported from model-driven software engineering (MDSE) and soft computing, and it is available through eGuider, a web portal for the enactment of computerised clinical practice guidelines and protocols.

MDSE provides the context and techniques to build new software artefacts based on models which conform to a specific metamodel. It also offers separation of concern, to disassociate medical from technological knowledge, thus allowing changes in one domain without affecting the other. The changes might include, for instance, the addition of new disorders to the DSS or new measures to the computation related to a disorder. Artificial intelligence and soft computing provide fuzzy logic to manage uncertainty and ambiguous situations. Fuzzy logic is embedded in an inference system to build a fuzzy inference system (FIS); specifically, a single-input rule modules connected zero-order Takagi-Sugeno FIS. The automatic creation of FISs is performed by the Aide-DS framework, which is capable of embedding the generated FISs in computerized clinical guidelines. It can also create a desktop application to execute the FIS. Technologically, it supports the addition of new target languages for the desktop applications and the inclusion of new ways of acquiring data.

DSScreening has been tested by comparing its predictions with the results of 152 real analyses from two groups: (1) NS samples and (2) clinical samples belonging to individuals of all ages with symptoms that do not necessarily correspond to an IEM. The system has reduced the time needed by 98.7% when compared to the interpretation time spent by laboratory professionals. Besides, it has correctly classified 100% of the NS samples and obtained an accuracy of 70% for samples belonging to individuals with clinical symptoms.

1. Introduction

Newborn screening (NS) is used to describe various tests that are done during the first few hours of the life of a newborn and, when properly timed and performed, have the potential for preventing severe health problems, including death (Therrell et al., 2015). Actually NS is intended to detect inborn errors of metabolism (IEMs) that can cause the sudden death of the newborn if they are not treated in time. On the contrary, if they are handled promptly, child mortality and morbidity are reduced (Hoffmann et al., 1996; Jones & Bennett, 2002; Sander et al., 2003), and the quality of life of infants and parents improves (Waisbren et al., 2003). NS is not a diagnosis procedure but a way of discarding a high percentage of the population minimizing the number of false-negatives (due to a low sensitivity) and false-positives (due to a low specificity). Therefore, a positive result does not imply that the newborn will necessarily develop the corre-

* Corresponding author.
E-mail addresses: unai.segundo@ehu.eus (U. Segundo), luisjose.aldamiz-echevarazuara@osakidetza.net (L. Aldámiz-Echevarría), javilo@ehu.eus (J. López-Cuadrado), david.buenestado@ehu.eus (D. Buenestado), fernando.andradelodeiro@osakidetza.net (F. Andrade), tomas.perez@ehu.eus (T.A. Pérez), raul.barrena@ehu.eus (R. Barrena), eduardo.gonzalezperez-yarza@osakidetza.net (E.G. Pérez-Yarza), jm.pikatza@ehu.eus (J.M. Pikatza).

http://dx.doi.org/10.1016/j.eswa.2017.02.022
0957-4174/

sponding disease. Whenever a positive is detected, supplementary diagnostic procedures are performed (either clinical, biochemical, or genetic) in order to definitively confirm the preliminary result and, in that case, treat the subject (Calderón López, Jiménez Parrilla, & Losada Martínez, 2008).

NS was first applied for the detection of Phenylketonuria (PKU) in the early sixties (Guthrie & Susie, 1963). Since then, this test has become part of the routine for newborn care and new techniques have been added to perform the procedure, for instance, enzyme-linked immunosorbent assay (Lequin, 2005), immunofluorescence (Fritschy & Härtig, 2001), and genetic testing or tandem mass spectrometry (MS/MS) (Millington, Kodo, Norwood, & Roe, 1990).

The medical outcome of NS practice is clear (Wilcken et al., 2009). According to (Wilcken, Rinaldo, & Matern, 2012), NS can detect over 50 IEMs, including disorders related to amino acids, fatty acid oxidation, and organic acids. However, not all disorders are always screened. Indeed, the set of IEMs included in the NS panels varies from country to country. The reasons for excluding an IEM may be different, such as particular national policies, ethical issues (Loeber et al., 2012); laboratory response times (Bodamer, Hoffmann, & Lindner, 2007); or the cost effectiveness of the test (Cipriano, Rupar, & Zaric, 2007).

Wilcken et al. (2012) state that screening programs are measured in terms of how they improve specificity (reducing the ratio of false-positives) without having a negative impact on sensitivity (the ratio of false-negatives). The response time is also critical, since some tests are discarded when they are too slow. The response time includes the time needed to perform the sample analysis using MS/MS equipment plus the time needed by laboratory professionals to evaluate the results and to create the corresponding report.

NS is usually performed to a batch of blood samples to be screened. Some control samples with known substances are inserted regularly with the aim of assuring calibration standards and also for quality reasons (Li & Tse, 2010). Once the results have been gathered, they are analysed manually by comparing the measures and several computed ratios to their established cut-off values. Determining whether all values are in their normality ranges is generally done by importing the data into Excel-like spreadsheets. Any change on the procedure, such as the inclusion of a new IEM to be screened, requires analysts to transfer their knowledge into that spreadsheet. It would also be problematic if there were a test that should be removed. Rather than erasing some measures or calculations (i.e., columns), analysts usually prefer not using them, as they could be used for another test.

Even with the help of some kind of software, NS is a time-consuming task. Thus, the development of artefacts such as decision support systems (DSSs) seems adequate because they can help laboratory professionals in their decisions. Moreover, it is highly advisable that these systems act as white box classification systems, which means that they do not only give a result, but also an explanation of their decisions.

Moreover, the knowledge about the medical evidence is not exact, so some tolerance to imprecision, partial truth or uncertainty is needed. In this sense, soft computing techniques (Kecman, 2001), such as fuzzy logic, machine learning by neural networks or support vector machines, or evolutionary computation are especially adequate for the medical domain. A fuzzy inference system (FIS) could also be a way to cope with inaccuracy of medical knowledge. Using a FIS, values coming from NS are fuzzyfied into linguistic values that a health professional can understand, and then they are given a value of certainty to their absolute value. Then the reasoning process occurs, and finally the FIS returns a conclusion (or a list of them). FISs are able to give their conclusions even when gathered data is incomplete. They usually present quality of information (QoI) indexes (Lima, Novais, Costa, Bulas-Cruz, & Neves, 2010) to measure the reliability of the results. The weights of the rules in a FIS can be defined using data mining techniques (Gadaras & Mikhailov, 2009) or by experts based on their experience and evidence from bibliography (Alayón, Robertson, Warfield, & Ruiz-Alzola, 2007). The former method is accuracy-oriented and the results obtained using it may lose interpretability (Zhou & Gan, 2008). The latter is more interpretable, but good prediction ability cannot always be guaranteed (Garibaldi, Zhou, Wang, John, & Ellis, 2012). Besides, a considerable effort is usually required in order for it to be changed or updated, because it integrates the medical knowledge with programming skills.

This paper presents DSScreening, an extendable DSS that uses a FIS as a core element. This FIS was created using the generative framework Aide-DS (Segundo et al., 2015). DSScreening is adaptable: the cut-off levels of the input variables and combinations of them can be changed. This is a way to accommodate to variations on the population measured. Each analyte (and its related ratios) has three linguistic values associated with it: *Low, Normal* and *High*. Their default cut-off levels (i.e. the absolute measures that define every linguistic value) have been defined, and the user is able to modify them if needed. Experts in the area have defined the weights assigned to the rules according to evidence in literature. The system is focused on helping to interpret the results of a MS/MS analysis by looking at the concentration of the relevant analytes for each IEM pattern. It is configured to display the most likely disorders (those with a probability equal to or higher than 50%) according to the measured values (if any) and explains why that conclusion has been reached.

Thanks to the separation of concerns (Ossher & Tarr, 2001) that model-driven software engineering (MDSE) provides, professionals can readjust DSScreening features, such as adding new disorders to be screened, substituting a test with a new one, or adjusting values to new medical evidence (that is, changing default values, adding new analytes to be taken into account or removing some of them). Indeed, regeneration can also be performed to adapt the system to technological changes without modifying the knowledge that models it. Thus, actions such as adding new sources of data (e.g., reading data provided by a new MS/MS device) or executing the DSS in another software platform can be carried out without much more effort.

The remainder of this paper is organized as follows. Section 2 will present research on the areas covered by this paper; Section 3 will provide specific information about the design and development of the implemented solution: the methods, how they collaborate to build DSScreening, and the samples used to measure its performance; Section 4 will present the results obtained for the samples presented in Section 3; Section 5 will discuss these results; and finally, Section 6 will state the conclusions of this work.

2. Related work

There is some evidence in literature about the use of DSSs for helping diagnose IEMs using soft computing techniques. For instance, Chen et al. (2013) present a framework based on machine learning classification as a tool to predict 3 different IEMs, taking the values of 35 analytes as input, in a web-based NS system. The framework contains a support vector machine as a decision model, which is a high-accuracy black-box classification system (i.e., it provides accurate results but without explanations). Marquardt et al. (2012) present a tool for matching input analyte measures with historical record values for specific disorders. It is the result of a worldwide initiative for standardising and improving NS based on multivariate pattern-recognition. In this system, users have to select the input analytes and the IEM they want to check. The system will automatically convert the input profile into