An Inference Engine for Estimating Outside States of Clinical Test Items

MASATO SAKATA, Advanced Telecommunications Research Institute
ZEYNEP YÜCEL, Advanced Telecommunications Research Institute
KAZUHIKO SHINOZAWA, Advanced Telecommunications Research Institute
NORIHIRO HAGITA, Advanced Telecommunications Research Institute
MICHITA IMAI, Advanced Telecommunications Research Institute and Keio University
MICHIKO FURUTANI, Tokyo Womens Medical University
RUMIKO MATSUOKA, Tokyo Womens Medical University and Toho University

Common periodical health check-ups include several clinical test items with affordable cost. However, these standard tests do not directly indicate signs of most lifestyle diseases. In order to detect such diseases, a number of additional specific clinical tests are required, which increase the cost of the health check-up. This study aims to enrich our understanding of the common health check-ups and proposes a way to estimate the signs of several life style diseases based on the standard tests in common examinations without performing any additional specific tests. In this manner, we enable a diagnostic process, where the physician may prefer to perform or avoid a costly test according to the estimation carried out through a set of common affordable tests. To that end, the relation between standard and specific test results is modeled with a multivariate kernel density estimate. The condition of the patient regarding a specific test is assessed following a Bayesian framework. Our results indicate that the proposed method achieves an overall estimation accuracy of 84%. In addition, an outstanding estimation accuracy is achieved for a subset of high-cost tests. Moreover, comparison with standard artificial intelligence methods suggests that our algorithm outperforms the conventional methods.

Our contributions are as follows: (i) promotion of affordable health check-ups, (ii) high estimation accuracy in certain tests, (iii) generalization capability due to ease of implementation on different platforms and institutions, (iv) flexibility to apply to various tests and potential to improve early detection rates.

Categories and Subject Descriptors: H.4.2 [**Types of Systems**]: Decision support; H.2.8 [**Database Applications**]: Data mining; I.2.1 [**Applications and Expert Systems**]: Medicine and science; J.3 [**Life and Medical Sciences**]: Medical information systems

General Terms: Clinical decision support, computational intelligence methodologies for health care, data analytics for clinical care.

Additional Key Words and Phrases: Data mining, health care, database, medical informatics, Bayesian inference.

This work was supported by JSPS KAKENHI Grant Number 24500545.
Author's addresses: M. Sakata, Yücel, Z., K. Shinozawa and N. Hagita, Advanced Telecommunications Research Institute, 2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288 Japan; M. Imai, Anzai-Imai Laboratory, Department of Information and Computer Science, Faculty of Science and Technology, Keio University 3-14-1, Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa-ken 223-8522 Japan; M. Furutani, The International Research and Educational Institute for Integrated Medical Sciences, Tokyo Womens Medical University, Tokyo, 162-8666, Japan and Department of Pediatric Cardiology, Tokyo Womens Medical University, Tokyo, Japan; R. Matsuoka International Research and Educational Institute for Integrated Medical Sciences, Tokyo Women's Medical University, Tokyo 162-8666, Japan and Department of Pediatrics, Faculty of Medicine, Toho University, Tokyo, Japan.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

ACM Reference Format:
Sakata, M., Yücel, Z., Shinozawa, K., Hagita, N., Imai, M., Furutani, M., Matsuoka, R., 2012. An Inference Engine for Estimating Outside States of Clinical Test Items ACM Trans. Manag. Inform. Syst. 0, 0, Article 00 (2012), 22 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION AND MOTIVATION

Health care industry is one of the earliest adopters of information technologies (IT) [Yasnoff et al. 2000]. In health care industry, IT has a broad application domain ranging from administration, to data collection, documentation and processing. An important application of IT in health care is clinical decision support systems.

Clinical decision support (CDS) systems are interactive tools, which are designed to assist the physicians in clinical decision tasks. CDS systems are used in various decision tasks in clinical medicine such as diagnosis of a disease, diagnostic process or patient management [Musen et al. 2006]. Among these various decision tasks, this study addresses in particular the application of CDS in diagnostic processes.

Diagnostic process refers to the selection of ordered tests or procedures and determining the value of the results relative to associated risks or costs [Musen et al. 2006]. Therefore, the decision task involves not only the analysis of patient data but also determination of the necessary tests and procedures adopting an integrated standpoint of risk and cost.

In general, CDS systems are composed of three main parts as illustrated in Figure 1 [Berner 2006]. The first part is the knowledge base, which is a set of known rules and associations such as drug-drug interactions or symptom-disease relations. The knowledge base is built based on the expert physician opinion and clinical practice guidelines [Garg et al. 2005]. The second part is called the inference engine, which contains the algorithms for combining the rules or associations in the knowledge base with actual patient data. Popular methods for building inference engines involve Bayesian networks, production rule systems and cognitive models of clinical reasoning. The combination of knowledge base and inference engine defines an expert system [Cowell et al. 2007]. Eventually, the third part, namely the communication mechanism, establishes an interaction interface between the system and the user, i.e. the physician. The communication mechanism enables inputting patient data into the system and reporting the output of the system to the user.

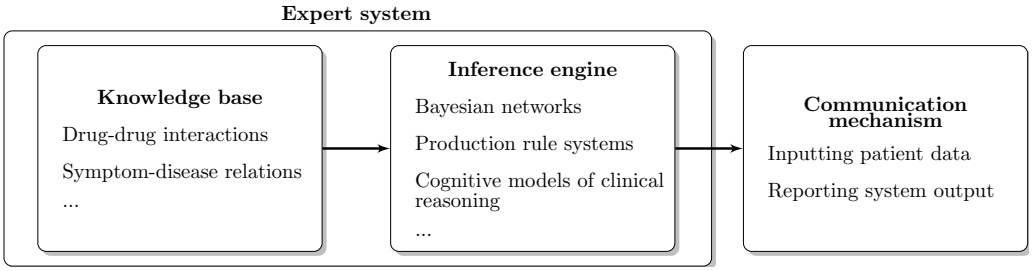

Fig. 1. Structural organization of a knowledge based clinical decision support system.

In this study we propose an inference engine to be employed by a CDS system in diagnostic processes. We benefit from the opinions of the expert physicians in building our knowledge base. Nevertheless, the main contribution of this study lies in the construction of the reasoning mechanism, which processes the health check-up data to

discover relationships between test values and patients' conditions. In addition to high estimation accuracy, we emphasize cost efficiency, generalization and flexibility in our design.

The outline of the paper is as follows. Section 2 provides a detailed overview of the relevant studies in literature. Section 3 describes the main components of the proposed method, introduces the terminology and explains the flow of the algorithm. Section 4 elaborates on the clinical database analyzed in this study. Subsequently, Sections 5 and 6 define the details of the estimation method. Section 7 presents the performance rates of the proposed method and of the standard pattern recognition methods together with a discussion on performance. Finally, Section 8 explains the distinguishing advantages and potential uses of the proposed method.

2. RELATED WORK

The improvement in quality of service and productivity introduced by the IT tools is ascertained by long term studies, which examine a relatively long time window in respect to the emergence and development span of IT tools [Menon et al. 2000]. A recent survey examining the evolution of CDS systems over five decades shows that recent CDS systems aim producing expert-level advisories, whereas early studies target rationalizing medicine by excelling in complex diagnostics tasks and outperforming clinicians [Peleg and Tu 2006]. Haux points out that health care will continue to take considerable advantage of the rapid and consistent improvements in IT in the future [Haux 2002]. A decade ago, Haux et al. described three requirements for effective use of IT in health care by the year 2013 as follows: (i) computerization of patient records with uniform terminology and standardized documentation, (ii) integration of knowledge into clinical work routine, and (iii) comprehensive use of patient data for clinical and epidemiological research [Haux et al. 2002]. Hitherto, certain improvements have been archived in terms of these points but obviously the operation in practice is far from being perfect.

Since CDS attracts more attention with the increasing use of computer systems in health care, numerous CDS systems have been proposed over the last few decades [Kaur and Wasan 2006; d'Aquin et al. 2006]. The key issue which these works deal with is processing the immense amount of clinical "data" so as to lead to "knowledge" [van Bemmel et al. 1997]. To that end, it is necessary to introduce models and methods to make sense of this data and attain knowledge out of it [Dwivedi et al. 2003]. In this respect, data mining methods are frequently used in CDS systems [Castellani and Castellani 2003; Fayyad et al. 1996].

Pawlak describes a rule base decision support tool, which can handle uncertainty and vagueness [Pawlak 1997]. To that end, a rough set approach is proposed. Nevertheless, this approach is not appropriate to medical framework, since medical probes are related often with nonlinear relationships and the rough set approach might easily fail in high dimensions. As opposed to exact and inflexible purely rule based methods, Kumar et al. proposed a hybrid approach using case-based and rules-based reasoning, which can handle problems with high complexity [Kumar et al. 2009]. In this manner, they propose making use of the tacit knowledge, which is subjective and nonstructured, as well as explicit knowledge, which is structured [Bose 2003]. Kong et al. provide an overview for such clinical decision support systems that has an uncertainty handling capability [Kong et al. 2008]. Rakus-Andersson et al. handle the problem from another point of view and propose an approximate reasoning based on the patient's clinical symptom levels and evaluate the operation risk [Rakus-Andersson and Jain 2009].

As well as the design of CDS systems, performance evaluation is a challenging topic too. Rahimi et al. provide a review on the evaluation studies of health information

systems in [Rahimi and Vimarlund 2007]. They point out that it is hard to quantify the performance of IT methods in health care. Besides, it is not possible to isolate the impact of IT from the other changes in the health care environment. Nevertheless, there are several common points such as quality of care, user and patience acceptance and satisfaction as well as financial effect, which are considered in most evaluation studies. DesRoches et al. consider these listed points in conducting a national survey in the US for evaluating electronic health records [DesRoches et al. 2008]. This survey reveals that although those physicians who often use such systems report high levels of satisfaction and increase in quality of service, the portion of these physicians is not large. Most health care professionals view financial cost of system implementation as the biggest barrier against extensive use of these tools. Additional problems are lack of training of technical staff and interoperability gap.

Kawamoto et al. examine seventy studies to identify the features which are critical in obtaining a successful CDS systems [Kawamoto et al. 2005]. They conclude that the success of those CDS systems has high correlation with four features, namely, automatic generation of decision support as part of the clinician work flow, delivery of decision support at the time and place of decision, delivery of actionable decisions, and computer integration.

Although most evaluation studies focus in identifying the features which make the system successful, some studies focus on failures and investigate the reasons. For instance, Littlejohns et al. describe the problems faced in installation of computerized integrated hospital information system in South Africa as opposed to implementation difficulties in highly developed countries such as US [Littlejohns et al. 2003]. They discuss on the failure of this project and conclude that the major reasons include social and professional culture of the health care organizations, cost of use education, underestimation of process complexity and installation duration, and the lack of learning from past failures. Heeks discusses the reasons of failure of health information systems at a larger scale and provides a review of studies reporting failures [Heeks 2006]. The reasons of failure are concluded to be gaps between design and reality, public and private sectors or countries.

3. OVERVIEW

We examine the existing CDS systems to learn the general trend, the desired attributes, and common factors in success as well as reasons of failures. In the light of these observations, we propose a new methodology, which distinguishes from the classical design framework in particular with its emphasis on cost efficiency and flexibility to be extended on different platforms and medical cases.

Our main goal is to provide a foresight to the physician regarding the condition of a patient (client) in terms of a clinical test or procedure, before it is actually performed. In the rest of this study, we use the term *test* or *test item* to describe these laboratory examinations or procedures, where clinical specimens are analyzed in order to get information about the health of a patient as pertaining to the diagnosis, treatment, or prevention of a disease. In light of the estimation carried out by the proposed algorithm, the physician may choose to order or skip these tests.

To that end, we examine the results of a set of clinical test items and try to estimate the *state* of the patient in terms of the test item of interest. The state of the patient refers his/her condition with respect to the *reference range*, which is a pair of lower and upper bounds of a test value. Reference ranges are used to estimate whether a test result's deviation from the mean is a result of a random variability or a result of an underlying disease or condition [Harris 1974; Noe 1985]. In this respect, this study distinguishes two sorts of state. Namely, provided that the test values of a patient are within the reference range, it is said that the deviation is random and thus the

patient has an *inside state*. Otherwise, the patient is said to have an *outside state*. We strongly emphasize that this study estimates the state of the patient with respect to the reference range as inside or outside, rather than estimating the exact value or range of values of the test item of interest.

In the rest of this study, we refer to the test item of interest, whose state is estimated, as the *estimated item*. The set of test items, which we base our estimation on, are referred as the *source items*. In order to uncover the relationship between the estimated item and source items, we consider the Cyber Integrated Medical Infrastructure (CIMI) database, which is an extensive collection of well-organized clinical test records [Abe et al. 2007; Shinozawa et al. 2009].

First, we expose the CIMI database to several pre-processing operations to enable a numerical computation and comparison. Pre-processing is composed of translation and alignment stages as illustrated in Figure 3. In translation, the non-numeric test results are converted to numeric values, whereas in alignment the test results are normalized to a common standard.

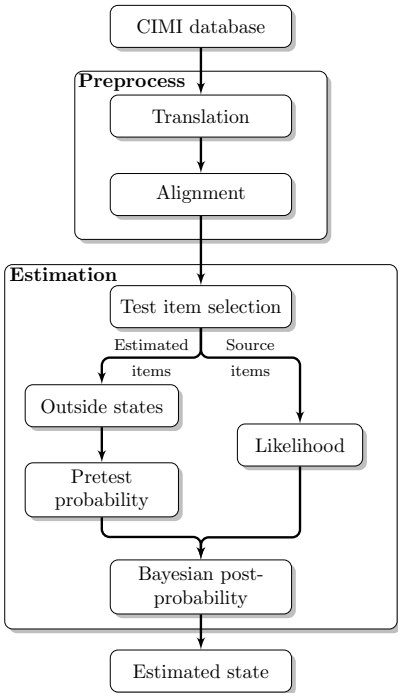

Fig. 2. Overview of the algorithm.

The estimation stage, which adopts a Bayesian approach, constitutes the essential element of the algorithm. Firstly, two sets of clinical test items are formed in the test item selection module, i.e. the estimated item and source items. In this stage, we impose the cost efficiency and flexibility conditions. Subsequently, the pretest probability is obtained using the previous values of the estimated item. The likelihood function is derived utilizing the values of the source items. Finally, a Bayesian post-probability is derived and the condition of the patient regarding the estimated item is assessed.

In what follows, we first elaborate on the CIMI database and then describe details of each stage of the algorithm.

4. CIMI DATABASE

We design and test our estimation method using the Cyber Integrated Medical Infrastructure database, which is developed by the International Research and Educational Institute for Integrated Medical Sciences [Abe et al. 2007; Shinozawa et al. 2009].

The CIMI Database involves medical examination values of a total of 122 test items recorded between August 2005 and October 2009. A total of 579 clients are monitored. Since ethics is an essential issue in health informatics, we paid utmost attention to volunteer participation and privacy [Goodman and Miller 2006]. The data is collected and stored with the consent of every individual client and is analyzed anonymously keeping the identification details confidential. Several specifications regarding the number of clients, average age, test frequency, and number of records are provided in Table I.

Table I. Specifications of the CIMI database.

	Male	Female	Total
Number of clients	260	319	579
Number of records	2311	1911	4222
Average age	56.7 ± 9.9	56.9 ± 11.1	56.8 ± 10.5
Test frequency (in months)	3.2 ± 1.2	3.3 ± 1.5	3.2 ± 1.3

Since this study aims designing a cost effective inference mechanism for a clinical decision support tool, we carry out a categorization of test items based on their cost. We categorize the test items into three classes: low-cost test items, medium-cost test items and high-cost test items. Table II lists the test items in low-cost and medium-cost categories. There are 48 test items in low-cost category. For the sake of brevity, we group the 48 low-cost test items into three, i.e. blood, urine and biochemical examination, and present the costs of these three groups together with the number of tests items in each group (7, 8, and 33, respectively). In addition, there are 6 test items in medium cost category. The remaining 68 test items in the database are regarded as high-cost test items and their total cost may exceed 1000 USD, whereas 9.25 USD is necessary to carry out the low-cost tests and 201 USD is required to perform medium-cost tests. The costs are given according to the Japanese health care system.

Table II. Categorization of test items according to their cost.

Category	Tests	Number of test items	Cost (USD)
Low-cost	Blood examination	7	3
	Urine examination	8	2.80
	Biochemical examination	33	3.45
Medium-cost	T-Cell CD2	1	50
	NSE	1	30
	Lipoprotein IDL	1	20
	Thymidine Kinase	1	45
	Apolipoprotein-B	1	16
	Pepsinogen I/II	1	40

5. PRE-PROCESSING

The pre-processing stage has two goals. The first one is to translate the test values which are provided as non-numeric values to numeric values. The second goal is to normalize the data recorded according to different standards.

5.1. Data translation

The results of most tests are provided as numeric values. However, several tests' results are expressed as non-numeric values. For instance, in serum tumor marker test, the existence of the Pepsinogen total factor is given in terms of semi-quantitative measurements, which define an approximation of the quantity of this substance without giving the exact amount. The results are expressed in terms of the degree of positivity or negativity, i.e. $(-)$, $(+-)$, $(+)$ or $(+2)$.

These non-numeric values are first quantified as -1, 0, 1, 2. However, such a discretization may easily lead to a singular covariance matrix as described in Section 6. In order to obtain a larger variation among the translated values, we make use of the definition of reference ranges. A well-accepted approach to define reference ranges is to assume a normal distribution for the test values and taking two standard deviations on either side of the mean concerning the reference group. In this manner, 95% of the samples are within the lower and upper limits, whereas 2.5% of the samples are below the lower limit and 2.5% of the samples are above the upper limit [Shultz et al. 1985].

For instance, Figure 3 demonstrates the distribution of Cholinesterase enzyme values for male patients in the CIMI database. Assuming a Gaussian distribution, the model illustrated with the blue dashed curve is obtained. The percentile of the samples below and above the reference ranges defined by the physicians are 1% and 2.4% for these test values. For the Gaussian distribution approximation, the percentile of the samples below the lower limit is 1.68% and the percentile of the samples above the upper limit is 1.22%.

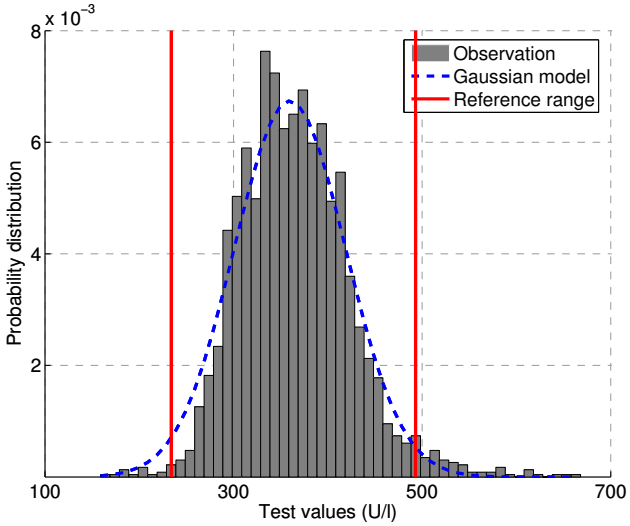

Fig. 3. The distribution of Cholinesterase enzyme measurements for male patients. The Gaussian distribution model and the reference ranges are illustrated with blue dashed line and a pair of red lines, respectively.

Similarly, we translate the discrete values making an additional assumption that each of them is a rounded approximation of a random variable coming from a normal distribution. In other words, a normal distribution is centered around each discrete value -1, 0, 1, and 2. The standard deviation of the normal distributions are picked according to the three-sigma rule[Pukelsheim 1994]. In general, a random variable coming from Gaussian distribution $\mathcal{N}(\mu, \sigma)$ is within $k\sigma$ around the expected value μ with a probability of $\mathrm{erf}(k/\sqrt{2})$. In our case, the centers of these distributions are assumed to be separated by six sigmas such that 99.7% of the samples fall within the corresponding range [Abramowitz et al. 1964]. Therefore, the condition that we enforce is tighter than the one usually used for reference ranges, in order to force the variables to be mapped into concerning intervals. Subsequently, for each discrete observation a random number is drawn from the concerning distribution.

An example of data translation concerning the Pepsinogen total factor test is illustrated in Figure 4. The blue lines indicate the original discretized values, whereas the red points denote the transformed numbers. As clearly seen in the figure, the proposed transformation scheme maps the discrete values almost always into separate bands.

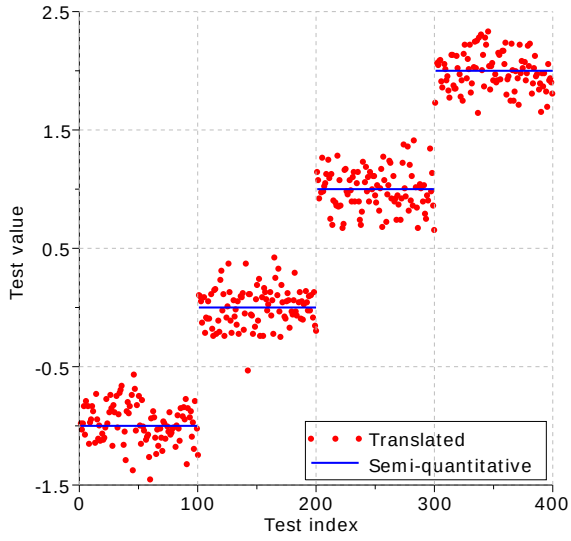

Fig. 4. The quantized semi-quantitative values of Pepsinogen total factor test are translated using the Gaussian distribution assumption. The figure illustrates 400 translations with 100 point concerning each discrete value.

Furthermore, in several other tests such as urinary sediment or mucus cord system tests, the results are expressed in terms of an interval instead of the absolute measured value, namely < 1, $1 \sim 4$, $5 \sim 9$, etc. Similar to the previous case, we translate these values by positioning a normal distribution at the center of each of these intervals, where the standard deviation is picked according to the three-sigma rule. Then, for each observation we draw a random number from the concerning distribution.

5.2. Data alignment

Reproducibility and standardization as well as overcoming the inter-operability gap are key issues which need to be addressed in design of clinical decision support systems [Trigo et al. 2012]. Provided that standardized tests and measures are employed, the operation range of the method is extended drastically enabling application across various institutions and platforms [Zaidi et al. 2002].

Therefore, we use commonly performed medical tests as a source basis. Nevertheless, due to the update of test standards, the range of values for several test items changed throughout the collection of the database and a direct comparison is not possible at all times. For instance, the results of liver bile and pancreas tests recorded before and after April 2006 are not directly comparable due to an update of test standards. In order to overcome this disparity, we propose a normalization scheme for the test values recorded according to different standards.

Let us assume the test values recorded according to the old standard (before April 2006) are denoted by y and the concerning reference range is between y_{min} and y_{max}. Let us further consider that according to the new standard, the reference range is shifted to x_{min} and x_{max}. The test values recorded according to the old standard are mapped to the new range using a linear interpolation. Namely, the mapped value x is obtained by the following equation,

$$x = (y - y_{min}) \frac{x_{max} - x_{min}}{y_{max} - y_{min}} + x_{min}.$$

6. STATE ESTIMATION

We aim to estimate whether a particular test item will have an inside or outside state before actually performing this test. For this purpose, we utilize the past values of this test item and the past and current values of a set of low-cost test items.

Let the estimated item and the set of source items be denoted by e and S, respectively. The set S is composed of individual source items s_n, i.e. $S = \{s_n\}$. Here, $1 \leq n \leq N$ as $\mathbf{n}\{S\} = N$, where $\mathbf{n}\{.\}$ gives the number of elements of a set.

Suppose that $x_c^e(\tau)$ stands for the value of test item e of client c at time τ. The state corresponding to this test value is determined by the following,

$$b_c^e(\tau) = \begin{cases} 0, & x_c^e \in I_e, \\ 1, & x_c^e \notin I_e, \end{cases}$$

where I_e stands for the reference interval of e, "0" denotes an inside state and "1" denotes an outside state. In the CIMI database, inside states are observed 473039 times, whereas outside states are observed 92709 times. Clearly inside states constitute the majority of observations.

We define another variable, namely *accumulated state*, which reflects the history of the patient's states over a given interval of length w. The accumulated state is denoted by $r_c^e(\tau)$ and is computed as the sum of the states $b_c^e(\tau)$ over the accumulation window, i.e. the last w time instants,

$$r_c^e(\tau) = \sum_{t=0}^{w-1} b_c^e(\tau - t). \tag{1}$$

Figure 5 presents an example for a series of test results and the corresponding states and accumulated states, where w is picked as 3. Obviously, due the causality property of Equation 1, we cannot compute values for time instants $\tau \leq w - 1$. Moreover, the accumulated state $r_c^e(\tau)$ can take values between 0 and w, since states are accumulated over w time instants. In this study, we model the probability density function (pdf) of

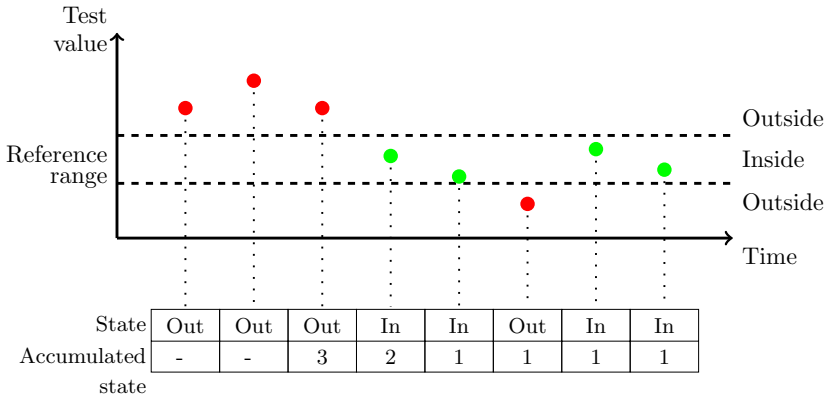

Fig. 5. Example of a computation of accumulated states for accumulation window $w = 3$.

each of the w possible accumulated states using a multivariate kernel density estimate.

To that end, we define a few additional variables for simplifying the notation. By listing the accumulated states of a test item e concerning client c between time $t = w$ and $t = \tau$, we obtain the vector of accumulated states, R_c^e,

$$R_c^e(\tau) = [r_c^e(\tau) \quad r_c^e(\tau - 1) \quad \cdots \quad r_c^e(w)]. \tag{2}$$

The variable R_c^e is used in computing the pretest probability of the accumulated state of the estimated test item in the following time instant.

In addition, we define another variable, namely $A_c^S(\tau)$, by smoothing the values in S,

$$A_c^S(\tau) = [a_c^{s_1}(\tau) \quad a_c^{s_1}(\tau) \quad \cdots \quad a_c^{s_N}(\tau)],$$

where the values $a_c^{s_i}$ are the smoothed values of $x_c^{s_i}$ by applying a moving average over each w consecutive samples,

$$a_c^{s_n}(\tau) = \frac{\sum_{t=0}^{w-1} x_c^{s_n}(\tau - t)}{w}.$$

The set of variables $A_c^S(t)$, $0 \leq t \leq \tau$, is used in determining the centers of the each of the w multivariate kernel density estimates.

Using the above defined variables, we estimate the probability that the accumulated state of a test item e at time $\tau + 1$ will be k. In other words, we do not directly estimate the state of e at $\tau + 1$, $b_c^e(\tau + 1)$. Instead, we estimate the value of its accumulated state $r_c^e(\tau + 1)$. Then, assuming that the states of the previous tests of e are provided, we can estimate its next state by comparing the estimated accumulated state $r_c^e(\tau + 1)$ and the sequence of past states $b_c^e(t)$, $\tau - w + 2 \leq t \leq \tau$.

In what follows, we describe the details of the estimation procedure. Section 6.1 elaborates on the selection source test items, whereas Sections 6.2 and 6.3 describe the computation of pretest probability and likelihood. Finally, Section 6.4 explains the computation of Bayesian post probability and the derivation of estimated state from the accumulated state.

6.1. Test item selection

In content developments of clinical decision support systems, a large number of health care professionals and information scientists are involved such as pharmacists, clinical analysts, software developers, physicians and administrators [Sittig et al. 2010]. In order to design and test the system in an effective manner, the interaction and collaboration between the members of this multi-disciplinary team has to be achieved. Recent studies show that an important part of the clinical communication space constitutes of direct interaction between clinicians [Coiera 2000]. Thus, in order to exploit the advantages of personal media, we work in close contact with physicians. In particular, the preliminary basis of this study is structured according to the advises of the physicians. Namely, test items which are used as a source for estimation are selected according to a rule base criteria advised by the physicians.

Initially, the low cost items listed in Table II are considered as a source set for estimation. Subsequently, the same estimation scheme is carried out once more, this time using both low-cost and medium-cost test items (See Table II). By comparing the two estimation performances, we assess the contribution of medium-cost test items in estimation performance.

6.2. Pretest probability

Let us assume that $P(r_c^e(\tau+1)=k)$ denotes the pretest probability that the accumulated state of a test item e concerning client c will be k at time $\tau+1$. In Bayesian sense, the pretest probability is derived based on the evidence provided by the past observations.

$$P\left(r_c^e(\tau+1)=k\right)=\frac{\mathbf{n}\left\{r_c^e=k|r_c^e\in R_c^e(\tau)\right\}}{\mathbf{n}\left\{R_c^e(\tau)\right\}},$$

where R_c^e is the accumulated state vector given by Equation 2.

6.3. Estimation of likelihood based on kernel density

Kernel density estimation is a non-parametric mean to estimate the pdf of a random variable [Parzen 1962]. Basically, the pdf is estimated by positioning a *kernel function* around each observation.

Figure 6 illustrates an example in one dimension. The instances of the random variable v are denoted by a red dot on the x-axis and a Gaussian kernel depicted by a blue dashed curve is centered at each instance of v. The variance of the Gaussian distribution, i.e. the *bandwidth* of the kernel function, plays a crucial role in the resulting estimate. The bandwidth, which is a free parameter, defines the amount and orientation of smoothing induced. Therefore, it controls the trade-off between the bias of the estimator and its variance. Kernel density estimation is preferred due to its smoothness and continuity properties, as opposed to the histogram based estimates which are discrete and discontinuous, but its bandwidth has to be adjusted carefully [Scott 1979].

We propose to compute the likelihood using the standard Gaussian kernel. Let G^{ek} denote the set of index pairs of clients and time instants for which the accumulated state of a test item e is k. Namely,

$$G^{ek}=\left\{c,\tau|r_c^e(\tau)=k\right\}.$$

The cardinality of the set G^{ek} is denoted by N^{ek}, i.e. $\mathbf{n}\left\{G^{ek}\right\}=N^{ek}$. We use G^{ek} to define the locations of kernel centers.

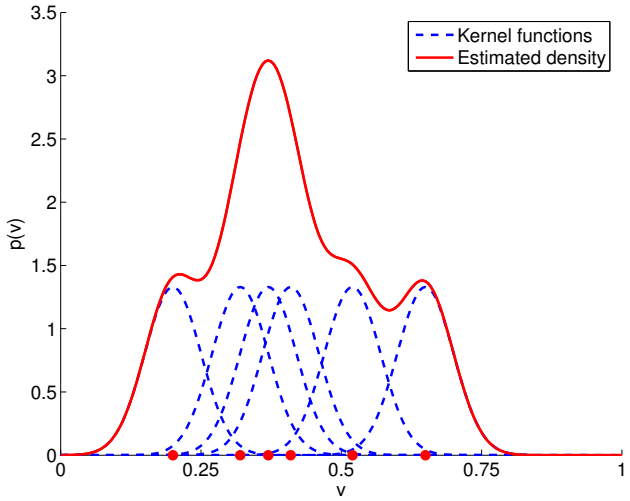

Fig. 6. Example for kernel density estimation in one dimension.

The likelihood of observing the source set S given $R_c^e(\tau)$, is given by

$$P\left(A_{c'}^S(\tau')|R_c^e(\tau)\right) = \frac{1}{h^{ek}N^{ek}} \sum_{c,\tau \in G^{ek}} K\left(\frac{A_{c'}^S(\tau') - A_c^S(\tau)}{h^{ek}}\right),$$

where $K(.)$ stands for the multidimensional kernel function and h^{ek} represents its bandwidth. For computing h^{ek}, we use Silverman's rule of thumb [Silverman 1986],

$$h^{ek} = \frac{1.06\Sigma}{\sqrt[5]{N^{ek}}},$$

where Σ is parametrized by a diagonal covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}.$$

Here, σ_i is the standard deviation of $\{x_c^{s_i}(\tau)\}, \forall c, \tau$.

6.4. Calculation of estimated state

The post-probability of observing each accumulated state is computed using the above defined pretest probability and likelihood functions in the Bayesian sense [Bishop et al. 2006]. Let p_k^e denote the post-probability that the accumulated state of a test item e at time $\tau + 1$ is k. The expected value of the accumulated state is computed as follows:

$$\widetilde{r_c^e}(\tau + 1) = \sum_{k=0}^{w-1} k p_k^e. \tag{3}$$

Obviously, the estimated value of accumulated state $\tilde{r_c^e}(\tau+1)$ is not necessarily discrete, whereas the actual values of accumulated states are always discrete as illustrated in

Figure 5. In order to obtain a discrete estimation, we apply the following operation:

$$\widetilde{\widetilde{r}}_c^e(\tau+1) = \begin{cases} \text{round}(\widetilde{r}_c^e(\tau+1)) & 0 \leq \widetilde{r}_c^e(\tau+1) \leq w-1, \\ w & w-1 < \widetilde{r}_c^e(\tau+1). \end{cases} \quad (4)$$

This is a basic rounding operation except for the positive bias towards outside states, where $\widetilde{r}_c^e(\tau+1)$ is very close to the accumulation window value w.

The estimated value of the next state $b_c^e(\tau+1)$ is derived from the estimated value of accumulated state $\widetilde{\widetilde{r}}_c^e(\tau+1)$ $r^{\sim e}_c$ in the following manner. We assume that client c had this test e at least $w-1$ times previously and that we know the states of these previous test results $b_c^e(\tau-t), \quad 0 \leq t \leq w-2$. The accumulated state estimate $\widetilde{\widetilde{r}}_c^e(\tau+1)$ can have a value, which is either equal to or 1 larger than the sum of previous states $\sum_{t=0}^{w-2} b_c^e(\tau-t)$. If it is equal, this indicates that the state in the next time instant $b_c^e(\tau+1)$ is 0, i.e. an inside state. If it is 1 larger, it means $b_c^e(\tau+1)$ is 1, i.e. an outside state. However, neither Equation 3 nor Equation 4 guarantees that this condition will be satisfied. Therefore, we modify this criteria as follows:

$$b_c^e(\tau+1) = \begin{cases} 0 & \widetilde{\widetilde{r}}_c^e(\tau+1) \leq \sum_{t=1}^{w-1} b_c^e(\tau-t), \\ 1 & \text{otherwise.} \end{cases}$$

The reason for the bias in Equation 4 is closely related this problem. As stated in Section 6, in the CIMI database inside states are observed more often than outside states. Therefore, the Bayesian approach favors inside states over outside states. However, in our diagnostic process framework, the costs of having different mistakes are not equal.

In medical informatics, these mistakes are quantified with *sensitivity* and *specificity*. Sensitivity is the fraction of disease cases that are correctly identified as disease. Specificity, on the other hand, is the non-disease cases that are identified as non-disease [Shin et al. 2006]. These terms are commonly used in medical informatics. Moreover, they correspond to equivalent terms in computer science or pattern recognition. Namely, sensitivity is often referred as true positive rate or recall rate, whereas specificity is referred as true negative rate [Provost et al. 1998].

Estimating a state which is actually outside, as inside, yields in overlooking the disease and thus a delay in diagnosis. Therefore, we give higher priority in detection of true positives than true negatives, or equivalently we can say that sensitivity is rated with higher importance than specificity. In case of Equation 4, estimating an accumulated state, which is actually w, as $w-1$ means that the future state is definitely estimated wrong. Moreover, a patient who always had outside states in the past, needs to be observed with more attention and precaution using this approach. Hence, we use the bias given in Equation 4 in order to detect accumulated states w as much as possible.

7. EVALUATION

We test our method by setting the accumulation window w to 3. In Table I, it is seen that the total number of test records is 4222. However, as mentioned in Section 6, we cannot make an estimation for $\tau \leq w$. By setting the accumulation window to 3, the number of time instants, for which we carry out an estimation, becomes 3080.

In each of these 3080 time instants, we perform a certain number of estimations. Namely, we perform an estimation for each test item, which is not in the source set S. Besides, for evaluating the proposed method in terms of cost, we propose testing the algorithm with two sorts of source sets, namely the low cost items (Low) and low and medium cost items (Low+Medium) (See Table II). In both cases, we estimate the states of the 68 high cost tests items.

7.1. Estimation performance of the proposed method

The performance of the proposed method in estimating states and accumulated states is presented in Table III. As seen in the table, accumulated states of "0" and "3" are estimated with higher accuracy than "1" and "2". For instance in (Low), detection rates of 87.28% and 76.30% are achieved in distinguishing "0" and "3", respectively, as opposed to 36.46% and 36.17% for "1" and "2". This is partially due to the fact that "0" and "3" indicate preservation of previous states. In other words, if the patient had outsides states in the past, he is likely to have an outside state at the following time instant, namely an accumulated state of "3". It is the similar case for inside states, i.e. "0". However, it is clearly more challenging to estimate state changes, which are observed when the accumulated state is "1" or "2".

Table III. Performance of the proposed method.

	Accumulated states(%)					States(%)		
Source set	0	1	2	3	Tot	Out	In	Tot
Low	87.28	36.46	36.17	76.30	80.32	60.87	87.38	84.14
Low+Medium	81.74	35.52	35.55	77.24	75.79	62.61	82.00	79.63

In addition, we investigate the failures in detection of accumulated states and illustrate what sort of confusions occur. The confusion matrix given in Table IV indicates that non-zero accumulated states are confused most often with "0". This means that most of the wrong estimations are underestimation, i.e. estimating outside states as inside. This is due the dominance of inside states in observations as stated in Section 6.

Table IV. Confusion table for (Low) and (Low+Medium).

		Estimated value of accumulated states							
		Low(%)				Low+Medium(%)			
		0	1	2	3	0	1	2	3
True value of accumulated states	0	87.28	0.31	0.05	12.36	81.74	0.43	0.06	17.77
	1	52.04	36.46	0.92	10.58	49.64	35.52	0.89	13.95
	2	42.94	9.40	36.17	11.49	40.81	8.88	35.55	14.77
	3	15.63	2.10	5.97	76.30	15.11	1.96	5.75	77.24

Furthermore, by looking at the detection rates of the proposed method on the right hand side of Table III, it is observed that inside states are estimated with higher accuracy than outside states, i.e. $87.38 > 60.87$ and $82.00 > 62.61$. Since specificity is the fraction of non-disease cases that are identified as non-disease and sensitivity is the fraction of disease cases that are correctly identified as disease as mentioned in Section 6.4, the proposed method has a very high specificity but sensitivity is not as

high as specificity . However, the overall performance rates in distinguishing outside and inside states appear to be satisfactory (84.14% and 79.63%), since inside states constitute the majority of the observations and thus have larger influence on the overall rate.

Additionally, comparing the two rows of Table III, it is observed that performance rates for (Low) and (Low+Medium) are not significantly different. Considering the cost of the low and medium cost items given in Table II, it is reasonable to perform the proposed estimation scheme using only low cost items for obtaining cost efficiency while keeping similar estimation performance.

Moreover, we investigate the test items whose states are estimated with lowest accuracy. As mentioned in Section 6.4, we give higher priority to detection of disease cases than non-disease cases, since overlooking a disease is more risky. Therefore, we sort the performance rates with respect to outside state detection rate.

Table V. Test items with lowest outside state estimation accuracy.

Low				Low+Medium			
Test name	Out (%)	In (%)	# of Out (out of 3080)	Test name	Out (%)	In (%)	# of Out (out of 3080)
Basophil	20.00	86.81	40	Basophil	25.00	80.92	40
CA 72-4	34.09	87.17	220	CA 72-4	38.64	81.78	220
Urobilinuria	38.89	86.70	38	Urobilinuria	39.47	81.10	38
SI	40.24	89.09	917	Neutrophil	40.60	80.59	133
Gamma seminoprotein	40.54	86.80	111	SI	41.98	84.97	917

Table V illustrates the test items, for which the proposed method fails to estimate the outside state most often. For the 5 test items with lowest performance, it is seen in Table V that detection rates of outside state change between 20.00% and 40.24%, which are significantly lower than the overall outside state detection rates of 60.87% and 62.61% given in Table III. One reason for these low rates is related to the number of outside state observations. As mentioned at the beginning of the section, by picking $w = 3$ we need to estimate states of 3080 time instants. Among these 3080 time instants, we need to have enough instances of outside state in order to capture the relation between the source items and these estimated items. However, it is seen from Table V that some items like Basophil and Urobilinuria have a total number of 40 and 38 instances of outside state, respectively. Obviously these numbers are quite low with respect to the total number of observations, 3080. Therefore, we are not able to model the relation between these items and the source set. However, certain test items such as SI, which have relatively high number of outside observations, are not estimated in a reliable manner with the proposed method.

There are several potential reasons for this. For instance, according to Linkov et al. CA 72-4 has a low intra-class consistency [Linkov et al. 2009]. In addition, Guadagni et al. argue that CA 72-4 has a sensitivity of approximately 40% to 50% opposed to a specificity of about 95%[Guadagni et al. 1995]. Moreover, some medical tests such as SI are interpreted not only individually but also together with the results of certain

other tests. In the future, these important combinations will be accounted for in our scheme.

In Table V, in addition to estimation rates of outside states, we also present the estimation rates of inside states. These rates ascertain that the algorithm does not have any classification bias. In other words, inside states of these tests are estimated with a similar performance to the one concerning the entire CIMI database. Namely, the rates in Table V change between 86.70% and 89.09% for (Low) and 80.59% and 84.97% for (Low+Medium), which are not significantly different than the values given in Table III (87.38% and 82.00%). Therefore, the proposed method detects inside states in a fair manner and the low detection accuracy of outside states is not due to any bias of the algorithm to classify all observations as inside.

Table VI. Test items with highest outside state estimation accuracy.

Low				Low+Medium			
Test name	Out (%)	In (%)	# of Out (out of 3080)	Test name	Out (%)	In (%)	# of Out (out of 3080)
H. pylori IgG	94.46	87.73	806	H. pylori IgG	94.79	83.11	806
Pepsinogen II	91.14	86.60	2789	Pepsinogen II	91.22	84.19	2789
Lipopotein	82.11	86.51	106	Gamma SM	84.21	80.95	19
RF	80.63	86.37	284	Lipoprotein	83.51	81.11	106
Gamma SM	78.95	87.00	19	RF	81.69	81.40	284

In addition to failure cases, we present the most successfully estimated test items in Table VI. As clearly seen comparing Tables VI and III, the detection rates of outside states of these items are significantly higher than the overall detection accuracy of outside states concerning the entire CIMI database. Besides, the detection rates of inside states in Table VI show that this high performance is not due to any bias in estimation. Additionally, the results are meaningful since the number of observations of outside states are significant with the exception of Gamma SM, which has only 19 outside state observations. Therefore, we can say that the proposed method can be used in estimating the states of the listed test items in Table VI in a reliable way. Furthermore, (Low) and (Low+Medium) do not have significant difference in performance also for the best estimated tests given in Table VI.

7.2. Estimation performance of the conventional methods

We compare the performance of the proposed method with several standard machine learning algorithms, namely linear discriminant classification (LDC), Quadratic discriminant classifier, K-nearest neighbor (KNN) classification and naive Bayes classification. The selection of classifiers are carried out in such a manner that several different classification approaches are tested. For instance, LDC uses a linear surface, whereas QDC uses a quadratic surface. Namely, LDC and QDC differ mainly in their assumption on class covariances. On the other hand, KNN is a non-parametric and non-linear classifier. Moreover, the naive Bayes classifier is a non-linear probabilistic classifier. In this manner, we employ linear or quadratic parametric classifiers as well

as non-parametric or probabilistic classifiers. The reader is referred to [Heijden et al. 2004] regarding the details of these classification algorithms.

Table VII. Performance of the conventional methods.

		Accumulated states(%)					States(%)		
Classifier	Source set	0	1	2	3	Tot	Out	In	Tot
LDC	Low	96.65	9.79	8.96	44.36	82.54	34.93	95.95	88.47
	Low+Medium	96.38	10.92	10.26	47.20	82.58	37.19	95.68	88.52
Naive Bayes	Low	90.28	14.76	14.24	41.25	77.70	37.09	89.63	83.20
	Low+Medium	90.32	15.72	14.74	44.42	78.05	39.29	89.65	83.48
KNN	Low	93.68	13.36	7.28	32.02	79.42	26.95	93.34	85.21
	Low+Medium	93.73	13.13	7.60	32.65	79.49	27.44	93.37	85.29
QDC	Low	71.60	16.41	13.04	34.39	62.18	31.47	72.65	67.61
	Low+Medium	69.84	16.11	12.50	34.27	60.70	30.83	71.11	66.17

First of all, by comparing each pair of rows in Table VII, it is observed that the performance rates regarding (Low) and (Low+Medium) do not have a significant difference.

It is observed that the accumulated state of “0” is estimated with high accuracy. However, “1” , “2” and “3” are estimated with much lower accuracy. Accumulated state of “3” is estimated with better accuracy than “1” and “2”, but its estimation rate is still very low compared to the proposed approach. Therefore, we can say that these methods have a bias in estimation and tend to classify most observations as “0”. In addition, we do not present the confusion tables for brevity’s sake, but as we examine them we observed that non-zero accumulated states are mostly confused with “0”.

By examining the right hand side of Table VII, it is observed that outside states are estimated with very low accuracy. Namely the alternative methods have an estimation varying between 27.44% and 39.29% for outside states. These numbers are significantly lower than the estimation rate of the proposed method for outside states (60.87% and 62.61%). Due to the bias of these methods to classify the observations as inside, the identification rate of inside states are quite high, namely between 95.95% and 71.11%. As a results of this bias and the majority of inside states in CIMI database, the overall recognition rate increases misleadingly.

Although the general detection rates of outside states concerning the entire CIMI database are low for these alternative methods, it is still possible that they are able to capture the distinguishing characteristics of outside and inside states for a very small number of test items. In other words, they might have very high performance for only a few tests and very low performance for most of the other tests. In that case, it would still be possible to employ these alternative methods for estimating this small subset. Thus we investigate whether the low rates in Table VII are due to a general incapability or an in-homogeneity in estimation performance between test items.

Therefore, we do the same thing, which we did in Table VI for our method, and depict the tests items which are estimated with highest accuracy. For the sake of brevity, we

do it for only one of the four alternative methods. We pick LDC, since it has the highest overall estimation performance as seen in Table VII. The five test items with highest outside state estimation accuracy regarding LDC are given in Table VIII.

Table VIII. Test items with highest outside state estimation accuracy for the LDC method.

Low				Low+Medium			
Test name	Out (%)	In (%)	# of Out (out of 3080)	Test name	Out (%)	In (%)	# of Out (out of 3080)
Pepsinogen II	97.24	21.99	2789	Pepsinogen II	96.20	50.52	2789
Apolipoprotein A1	70.04	77.59	1442	Apolipoprotein A1	70.60	70.82	1442
HDL	61.26	88.91	888	HDL	62.73	86.46	888
HbA1c	49.15	97.90	177	H. pylori IgG	62.53	92.92	806
Albumin fraction	48.51	97.20	505	HbA1c	50.85	97.76	177

Clearly, the outside states concerning Pepsinogen II are estimated with a very high accuracy for both (Low) and (Low+Medium). However, note the estimation accuracy of inside state for these cases, which is very low. This indicates that LDC has a tendency to classify most observations of Pepsinogen II as outside. This is due to the high observation frequency of outside states in Pepsinogen II test (2789 out of 3080). Apolipoprotein A1 suffers from the same drawback. Although the estimation rates are more or less fair for (Low+Medium), they are still not as good as the rates of the proposed method (See Table VI). Besides, considering the 201 USD additional cost of carrying out (Low+Medium) scheme instead of (Low), the performance rates of these alternative methods are clearly not satisfactory.

Table IX. Test items which are estimated with highest outside state accuracy with proposed method are estimated with lower accuracies by LDC.

Low				Low+Medium			
Test name	Out (%)	In (%)	# of Out (out of 3080)	Test name	Out (%)	In (%)	# of Out (out of 3080)
H. pylori IgG	22.08	92.57	806	H. pylori IgG	62.53	92.92	806
Pepsinogen II	97.24	21.99	2789	Pepsinogen II	96.20	50.52	2789
Lipopotein	9.82	96.89	106	Gamma SM	0	98.92	19
RF	1.41	97.00	284	Lipoprotein	23.16	96.60	106
Gamma SM	0	98.76	19	RF	1.41	97.57	284

Moreover, in Table IX exactly the same set of test items appearing in Table VI are handled. In this manner, we demonstrate the performance of LDC for those tests items, which are estimated with best accuracies with the proposed method. It is conclude that these particular tests are estimated with very poor accuracies by LDC. In addition, the bias of LDC in favouring the frequently appearing states is confirmed once again. For instance, H. Pylori IgG has a higher number of inside states than outside states (806 out of 3080). The effect of this difference is clearly observed in estimation accuracies (22.08% vs 92.57%). Similarly, the dominance of outside states for Pepsinogen II (2789 out of 3080) is reflected in the estimation accuracy as a strong bias towards outside states(22.08% vs 92.57%).

8. DISCUSSION

This study proposes an inference mechanism to be used in diagnostic process tasks of clinical decision support systems, which distinguishes from the classical design framework with its high accuracy, emphasis on cost efficiency as well as generalization capability and flexibility to be extended on different platforms. In what follows, we discuss in detail on each of these advantages.

The proposed inference framework proves to achieve a very high specificity. Overall sensitivity is not as high as specificity. Nevertheless, it is shown that it is still much better than those of standard machine learning algorithms. In addition, it is not reasonable to expect estimation of all the tests using the same set of standard test items. Therefore, we examine the test items which are estimated with highest and lowest accuracy and demonstrate that the proposed method can be used confidently in estimating the states of a set of high-cost test items (See Table VI).

In addition to achieving high accuracy in certain tests, our method emphasizes cost efficiency in diagnostics procedure. Namely, with the proposed system, the routine test items, which have a relatively low cost, can be used to determine the condition of a patient with respect to another test item with a high cost. Furthermore, by employing two sorts of source sets, namely (Low) and (Low+Medium), we ascertain that the additional test items with medium cost do not contribute to the improvement of estimation accuracy to a significant degree. This suggests that a source set composed of low cost test items provides a high enough accuracy with a reasonable cost.

Moreover, provided that the proposed inference mechanism is integrated into a CDS, the direct link established between the estimation mechanism and the physicians offers a profound integration of financial standpoint into the medical diagnosis scheme. Since physicians make the most important decisions such as hospital admittance, ordering tests and procedures, and prescription of medications, this link has the potential to control a significant portion of medical care costs [Cohen et al. 1982; Enthoven 1980].

Furthermore, the method has high generalization capabilities. Generalization refers to application of the proposed scheme on different platforms and in different institutions. Since the source test constitutes of common health check-up tests, the implementation of the proposed inference framework is quite straightforward. Namely, it does not require any changes in the practical treatment procedure. Therefore, it is possible to apply it in different environments without changing the medical conventions.

Besides, the estimation mechanism is flexible to apply into different medical cases. Here, medical case refers to the treatment of a particular patient, whereas flexibility refers to the adaptation of the method to different scenarios with varying potential diseases. Once the decision framework is established, it can be applied to any medical case or for any disease in the database. In this manner, the physician has the possibility to investigate indications of not only a disease of first concern but several other diseases in the database without additional effort or cost. For instance, the proposed

clinical decision support system employs a set of generic test items to estimate oncology tests. This means that even thought the physician considers cancer to be a slight possibility, he/she still has a possibility to investigate the likeliness. Therefore, the commonly performed tests might easily be used to estimate the condition of the patient in respect of any unanticipated disease without consuming additional resources. This process might seem redundant in most cases but keeping in mind that the early detection and treatment is of utmost importance in oncology along with many other medical fields, the effort for estimating the result of these tests is minimal considering the possible benefits.

Furthermore, the proposed method has the potential to leverage public health informatics promoting the health of populations [Yasnoff et al. 2000]. Namely, the subset of low-cost medical tests are applied for most patients taking a medical check-up, which enables monitoring the condition of a community as well as single individuals with respect to a costly test.

9. CONCLUSIONS

This study proposes an effective analysis for the "information rich" yet "knowledge poor" health care data in a cost effective manner. We utilize a set of standard test items in common health check-ups to estimate the future states of a set of costly medical tests. The relation between test states is modeled through a multivariate kernel density estimate in a Bayesian framework. Our results demonstrate that the proposed method achieves significant performance in an extensive database. Moreover, we discover that a subset of high-cost tests are estimated with outstanding accuracy and thus the proposed algorithm can be used for estimating those tests in a reliable manner. In addition to high estimation accuracy and cost effectiveness, our algorithm bears a number of potential benefits. First of all, with the proposed system the physician might investigate the condition of a patient with respect to a slightly anticipated test without any additional effort or cost. This might enhance early detection of several diseases and improve health care quality. Besides, a source set composed of commonly performed tests enables monitoring the condition of not only individual patients but also a community. Moreover, the system is relatively easy to integrate into the current health care platforms, due to the common application of the source tests.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Francesco Zanlungo for his valuable discussion.

REFERENCES

ABE, A., HAGITA, N., FURUTANI, M., FURUTANI, Y., AND MATSUOKA, R. 2007. Possibility of integrated data mining of clinical data. *Data Science Journal 6,* 1, 104–115.

ABRAMOWITZ, M. E. ET AL. 1964. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Vol. 55. Courier Dover Publications.

BERNER, E. 2006. *Clinical decision support systems: theory and practice*. Springer.

BISHOP, C. ET AL. 2006. *Pattern recognition and machine learning*. Vol. 4. springer New York.

BOSE, R. 2003. Knowledge management-enabled health care management systems: capabilities, infrastructure, and decision-support. *Expert Systems with Applications 24,* 1, 59–71.

CASTELLANI, B. AND CASTELLANI, J. 2003. Data mining: qualitative analysis with health informatics data. *Qualitative Health Research 13,* 7, 1005–1018.

COHEN, D., JONES, P., LITTENBERG, B., AND NEUHAUSER, D. 1982. Does cost information availability reduce physician test usage?: A randomized clinical trial with unexpected findings. *Medical care*, 286–292.

COIERA, E. 2000. When conversation is better than computation. *Journal of the American Medical Informatics Association 7,* 3, 277–286.

COWELL, R., DAWID, P., LAURITZEN, S., AND SPIEGELHALTER, D. 2007. *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer.

D'AQUIN, M., LIEBER, J., AND NAPOLI, A. 2006. Adaptation knowledge acquisition: A case study for case-based decision support in oncology. *Computational intelligence 22,* 3-4, 161–176.

DESROCHES, C., CAMPBELL, E., RAO, S., DONELAN, K., FERRIS, T., JHA, A., KAUSHAL, R., LEVY, D., ROSENBAUM, S., SHIELDS, A., ET AL. 2008. Electronic health records in ambulatory carea national survey of physicians. *New England Journal of Medicine 359,* 1, 50–60.

DWIVEDI, A., BALI, R., NAGUIB, R., AND NASSAR, N. 2003. Clinical knowledge management for healthcare. In *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*. IEEE, 157–160.

ENTHOVEN, A. 1980. *Health Plan: The Practical Solution to the Soaring Cost of Medical Care*. Beard Books Incorporated.

FAYYAD, U., PIATETSKY-SHAPIRO, G., AND SMYTH, P. 1996. From data mining to knowledge discovery in databases. *AI magazine 17,* 3, 37.

GARG, A., ADHIKARI, N., MCDONALD, H., ROSAS-ARELLANO, M., DEVEREAUX, P., BEYENE, J., SAM, J., AND HAYNES, R. 2005. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA: the journal of the American Medical Association 293,* 10, 1223–1238.

GOODMAN, K. AND MILLER, R. 2006. Ethics and health informatics: Users, standards, and outcomes. *Biomedical Informatics*, 379–402.

GUADAGNI, F., ROSELLI, M., COSIMELLI, M., FERRONI, P., SPILA, A., CAVALIERE, F., CASALDI, V., WAPPNER, G., ABBOLITO, M. R., GREINER, J. W., ET AL. 1995. Ca 72-4 serum marker-a new tool in the management of carcinoma patients. *Cancer investigation 13,* 2, 227–238.

HARRIS, E. 1974. Effects of intra-and interindividual variation on the appropriate use of normal ranges. *Clinical Chemistry 20,* 12, 1535–1542.

HAUX, R. 2002. Health care in the information society: What should be the role of medical informatics? *Methods of information in medicine 41,* 1, 31–35.

HAUX, R., AMMENWERTH, E., HERZOG, W., KNAUP, P., ET AL. 2002. Health care in the information society. a prognosis for the year 2013. *International Journal of Medical Informatics 66,* 1, 3–22.

HEEKS, R. 2006. Health information systems: Failure, success and improvisation. *International journal of medical informatics 75,* 2, 125–137.

HEIJDEN, F., DUIN, R., RIDDER, D., AND TAX, D. 2004. Classification, parameter estimation and state estimation: an engineering approach using matlab. *Cornwall, UK: John Wiley & Sons. p*, 183–4.

KAUR, H. AND WASAN, S. 2006. Empirical study on applications of data mining techniques in healthcare. *Journal of Computer Science 2,* 2, 194–200.

KAWAMOTO, K., HOULIHAN, C., BALAS, E., AND LOBACH, D. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj 330,* 7494, 765.

KONG, G., XU, D., AND YANG, J. 2008. Clinical decision support systems: a review on knowledge representation and inference under uncertainties. *International Journal of Computational Intelligence Systems 1,* 2, 159–167.

KUMAR, K., SINGH, Y., AND SANYAL, S. 2009. Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in icu. *Expert Systems with Applications 36,* 1, 65–71.

LINKOV, F., GU, Y., ARSLAN, A. A., LIU, M., SHORE, R. E., VELIKOKHATNAYA, L., KOENIG, K. L., TONIOLO, P., MARRANGONI, A., YURKOVETSKY, Z., ET AL. 2009. Reliability of tumor markers, chemokines, and metastasis-related molecules in serum. *Eur Cytokine Netw 20,* 1, 21–26.

LITTLEJOHNS, P., WYATT, J., AND GARVICAN, L. 2003. Evaluating computerised health information systems: hard lessons still to be learnt. *Bmj 326,* 7394, 860–863.

MENON, N., LEE, B., AND ELDENBURG, L. 2000. Productivity of information systems in the healthcare industry. *Information Systems Research 11,* 1, 83–92.

MUSEN, M., SHAHAR, Y., AND SHORTLIFFE, E. 2006. Clinical decision-support systems. *Biomedical informatics*, 698–736.

NOE, D. 1985. *The logic of laboratory medicine*. Urban & Schwarzenberg.

PARZEN, E. 1962. On estimation of a probability density function and mode. *The annals of mathematical statistics 33,* 3, 1065–1076.

PAWLAK, Z. 1997. Rough set approach to knowledge-based decision support. *European journal of operational research 99,* 1, 48–57.

PELEG, M. AND TU, S. 2006. Decision support, knowledge representation and management in medicine. *Yearb Med Inform 45*, 72–80.

PROVOST, F., FAWCETT, T., AND KOHAVI, R. 1998. The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the fifteenth international conference on machine learning*. Vol. 445.

PUKELSHEIM, F. 1994. The three sigma rule. *The American Statistician 48,* 2, 88–91.

RAHIMI, B. AND VIMARLUND, V. 2007. Methods to evaluate health information systems in healthcare settings: a literature review. *Journal of medical systems 31,* 5, 397–432.

RAKUS-ANDERSSON, E. AND JAIN, L. 2009. Computational intelligence in medical decisions making. *Recent Advances in Decision Making*, 145–159.

SCOTT, D. 1979. On optimal and data-based histograms. *Biometrika 66,* 3, 605–610.

SHIN, H., MARKEY, M., ET AL. 2006. A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples. *Journal of Biomedical Informatics 39,* 2, 227–248.

SHINOZAWA, K., HAGITA, N., FURUTANI, M., AND MATSUOKA, R. 2009. A data mining method for finding hidden relationship in blood and urine examination items for health check. *Advances in Data Mining. Applications and Theoretical Aspects*, 44–50.

SHULTZ, E., WILLARD, K., RICH, S., CONNELLY, D., AND CRITCHFIELD, G. 1985. Improved reference-interval estimation. *Clinical chemistry 31,* 12, 1974–1978.

SILVERMAN, B. 1986. *Density estimation for statistics and data analysis*. Vol. 26. Chapman & Hall/CRC.

SITTIG, D., WRIGHT, A., SIMONAITIS, L., CARPENTER, J., ALLEN, G., DOEBBELING, B., SIRAJUDDIN, A., ASH, J., AND MIDDLETON, B. 2010. The state of the art in clinical knowledge management: an inventory of tools and techniques. *International journal of medical informatics 79,* 1, 44–57.

TRIGO, J., MARTÍNEZ, I., ALESANCO, A., KOLLMANN, A., ESCAYOLA, J., HAYN, D., SCHREIER, G., AND GARCÍA, J. 2012. An integrated healthcare information system for end-to-end standardized exchange and homogeneous management of digital ecg formats. *Information Technology in Biomedicine, IEEE Transactions on 16,* 4, 518–529.

VAN BEMMEL, J., MUSEN, M., AND HELDER, J. 1997. *Handbook of medical informatics*. Bohn Stafleu Van Loghum Houten.

YASNOFF, W., O CARROLL, P., KOO, D., LINKINS, R., AND KILBOURNE, E. 2000. Public health informatics: improving and transforming public health in the information age. *Journal of Public Health Management and Practice 6,* 6, 67–75.

ZAIDI, S., ABIDI, S., AND MANICKAM, S. 2002. Distributed data mining from heterogeneous healthcare data repositories: towards an intelligent agent-based framework. In *Computer-Based Medical Systems, 2002.(CBMS 2002). Proceedings of the 15th IEEE Symposium on*. IEEE, 339–342.

Received November 2012; revised XX 2012; accepted XX 2012