Journal of Information Processing Vol.20 No.1 327–339 (Jan. 2012)

[DOI: 10.2197/ipsjjip.vol.20.327]

Regular Paper

Discovery of Patterns to Improve Usability of Electronic Health Record Systems

Shelly Sachdeva[1,a)] Aastha Madaan[1] Subhash Bhalla[1]

Received: February 7, 2011, Accepted: September 12, 2011

Abstract: A majority of research efforts in the domain of Electronic Health Records concentrate on standardization and related issues. The earlier forms of medical records did not permit a high level of exchange, interoperability or extensive search and querying. Recent research has focused on the development of open standards for life-time long health record archives for individual patients. This facilitates the extensive use of data mining and querying techniques for the analysis. These efforts can increase the depth and the extent of the utilization of patient data. For example, association analysis can be used to identify common features among disparate patients to check whether diagnoses or procedures are effective. Pattern discovery techniques can also be used to create the census reports and generate a meaningful visualization of summary data at hospitals. For handling the large volume of data, there is a need to focus on improving the usability. The current study proposes a model for the development of EHR support systems. It aims to capture the health worker's needs in a scientific way, on a continuous basis. The proposal has been evaluated for the accuracy of knowledge discovery to improve the usability.

Keywords: data quality, electronic health record, human-computer interaction, interoperability, usability

1. Introduction

Electronic Health Record (EHR) systems can track medical errors, save time and provide better care. However, users often face the problem of loss of productivity and steep learning curves. These may eventually lead to failures. Secondly, the evolution of EHRs and the increase in complexity in the medical domain is an ongoing phenomenon. An improved user-centered design (UCD) approach is needed to involve end-users throughout the development process. Thus, the technology supported tasks must be easy to operate, and must offer a higher level of response. This paper proposes a model. The model captures the user behavior, task requirements, and the work flow. It's aim is to utilize the development process of EHR systems to capture user's needs. It facilitates the improvement of the usability of EHR systems.

1.1 Role of Electronic Health Records

EHRs refer to the lifelong collection of the medical record of a patient. The Healthcare Information and Management Systems Society (HIMSS) defines an EHR system, as the clinical information system owned and operated by a healthcare delivery organization, that serves as the legal record of patient encounters [2]. Further, the EHR evolution has led to the evolution of Personal Health Records (PHRs). In case of PHRs, a software application is used by individuals to record their personal knowledge about their health and the health of their dependents. There are web-based PHR portals – namely, Google Health [12] and Microsoft Health Vault [13]. These aim to store and allow access to the EHR data from a variety of perspectives. An example of an existing healthcare organization that utilizes EHRs is shown in **Fig. 1**.

A hospital maintains an archive of the EHRs of a patient as a part of its day-to-day activity. In comparison to PHRs, the EHR repositories at the hospitals, provide multiple additional functions. These facilitate the epidemic prevention studies and facilitate national planning using a life-long record for a population. Furthermore, these support shifting towards an integrated care for the healthcare, the medical care and social-welfare activities. However new questions with respect to larger systems and the volume of data may arise.

Hospitals and health agencies exchange EHR data through standards such as HL7 (Health Level 7) [17] and DICOM (Digital Imaging and Communications in Medicine) [18]. The standards such as, CEN EN 13606 [20] and openEHR [20] define the unit of exchange (documents) as EHR_Extract (expressed in XML). In order to have interoperable exchanges, the EHRs use industry standards promoted by the Integrating Healthcare Enterprise (IHE) [19]. There are many challenges concerning the usability of EHRs. At the present moment, the EHRs are not designed to be user centric and pose problems caused by information overload. The usability study of EHR systems needs to focus on the flexible navigation, the personalization and the customization, enabling data variations and visualizations [1].

1.2 Contents of Electronic Health Records

Standardized EHRs of a patient have an extensive and maximal structure that may include data from about 100–200 parameters, such as the temperature, the blood-pressure and the body mass index [20]. Individual parameters will have their own contents.

1 Graduate Department of Computer and Information Systems, University of Aizu, Aizu-wakamatsu, Fukushima 965–8580, Japan

a) d8111107@u-aizu.ac.jp

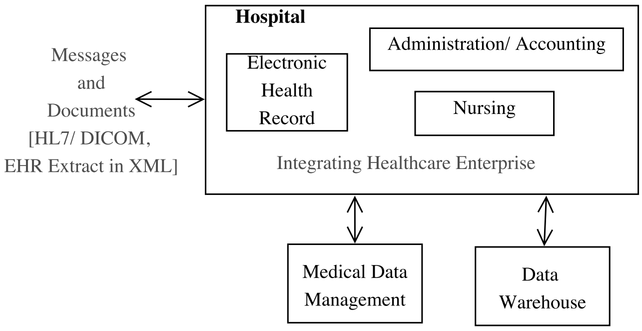


Fig. 1 A Healthcare organization.

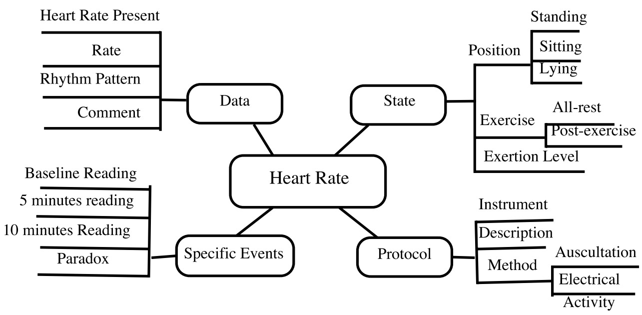


Fig. 2 Heart rate archteype.

Each contains an item, such as 'data' (e.g., captured for a heart rate observation). It offers the complete knowledge about a clinical context, (i.e., attributes of data), 'state' (context for interpretation of data), and 'protocol' (information regarding gathering of data) (**Fig. 2**). There is a need for a standardized communication format and protocol, since a patient's health information is shared in a multi-disciplinary (shared care) environment. Thus, the development and the adoption of national and international standards for EHR interoperability is likely to make a large volume of data available for diagnostics and research. The openEHR foundation (and other standards organisations, such as HL7 and CEN13606) are making efforts for enabling Information and Communication Technology to effectively support healthcare, medical research and related areas (www.openehr.org). Contents within EHRs are being defined to support medical and healthcare applications. The dual model approach, proposed by openEHR has been accepted by all EHR standards (HL7, openEHR and ISO 13606). It includes reference modeling (information modeling) and content modeling. Reference model (RM) contain a stable and fundamental structure. It represents the generic structures of the components of health record information, at the storage level. Content models, on the other hand, cover the informational aspects which are not as stable due to the variability and the high rate of change of the domain knowledge (i.e., the formal description of the physical examination or the prescription). The first layer consisting of reference model contains a generic class 'ACTOR'. In the second layer, through the use of content models (Archetypes), this generic class can be extended to be 'Doctor', 'Nurse' or 'Insurer'. An archetype is an agreed formal and interoperable specification of a re-usable clinical data set which underpins an electronic health record (EHR), capturing as much information about a particular and discrete clinical concept as possible. An example of a simple archetype is the heart rate (Fig. 2), which can be used in multiple places, wherever it is required within an EHR.

1.3 Human Interface for End Users

Considering the human factors, physicians use complex terms (or a foreign terminology) to describe diagnoses. They have task differences such as, the pace of work, the delegation, the data entry and different perspectives on health. They usually complain of forced changes to establish successful work flows, long training times, and the fact that, they have to deal with complex tasks and functionalities. In this regard, they need the support of an efficient information system, with a high level of usability on an ongoing basis [1]. Further, the desired functionality and data entry perspectives may vary for various categories of users.

At the system side, the navigation scheme must be designed to meet the needs of a medical professional. A user should be able to access the information in a few steps. Different specialists should be able to personalize the interface to carry out their task flows and data entries. In order to ease the work of physicians, the use of defaults can be made. Thus, if the physician selects a running nose as a complaint, the interface must automatically present the defaults for laboratories, medicine and notes.

Additionally, to save time, the interface should support the rapid switching between patients and should also allow the physician to inherit the login used by the nurse who noted the vital details and recorded patient's data and vice-versa. This can facilitate the completion of the administrative aspects of the visit. For data entry and retrieval, templates can be used that are thoroughly tested by user [1]. Similarly, to deal intelligently with problems regarding human physiology, the information should be displayed in formats that can be integrated and can be manipulated. In order to meet the complex set of needs, research techniques are needed for the analysis and the study.

1.4 Role of Pattern Analysis and Discovery

Pattern discovery is the process of finding an interesting knowledge from large amounts of data stored in information repositories. The interesting patterns are detected, presented to the user, and stored as new knowledge in the knowledge base. These data mining tasks can be descriptive or predictive. The descriptive tasks allow the users to characterize the general properties of the data. Similarly, the predictive tasks perform inferences on the current data to make predictions [14]. The knowledge can be used for the decision making process and to forecast effects of those decisions. These can be effectively designed and used to improve the usability of the EHR systems. The uniqueness of medical data mining is due to the heterogeneity of medical data, ethical, legal and social issues, the statistical philosophy and the special status of medicine [23]. The benefits can be highlighted through the examples.

- Association Rules: Techniques in association rule mining can find medications that frequently co-occur in EHRs.
- Classification and Prediction: Through a classification technique
  (1) The system can decide if there is a need for an MRI exam for a given patient considering the similarities

among EHRs;

(2) The prediction technique could help predict the length of hospitalization for a specific patient after hospitalization;

(3) It can be seen whether a hip replacement treatment is doing well (by observing after the discharge, whether the patient has to come back multiple times, or whether he has an infection); and

(4) It can be seen whether or not a bare metal stint is better considering EHRs of a population. In general, this amounts to looking at the effectiveness of a stint. It is a high-cost item, and the hospital can see if one brand is doing better than another in a population.

- Clustering: Mining techniques in clustering can form groups based on the selected properties.
- Evolution: Pattern discovery in evolution aims to find any interesting pattern in medication time series.

In addition to the medical records, other related data repositories can be explored for a similar purpose. For example, data about drugs, proteins, genes and EHR related bio-medical entities can be mined. Biomedical data mining is becoming a common component in the biomedical informatics. Recent research emphasizes the discovery of important patterns or irregularities in the huge volumes of biomedical data [27]. The following examples highlight the use of pattern discovery and data mining in the biomedical domain:

(1) Bayesian network-based method for modeling the survival of patients in hospitals, which allows the expected cost to be estimated for the patients accumulated duration of time in care [27].

(2) Simulated annealing for the bi-clustering of gene expression data using a stochastic search technique [27].

(3) Model-based approach to extract the high-level length of stay patterns of residents in long-term care [27].

Hritidis [16] gave various examples of mining applications for EHR, such as,

(1) The analysis of real-time EHR data to monitor vital signs and generate alerts.

(2) Decision tree techniques to develop and validate a measure of risk stratification for in-hospital mortality due to accurate break down of heart failure.

(3) Association rule mining [15] to create rules about the co-occurrence of pathology tests in data collected in the health insurance industry (specifically, an episode (claims) database).

Data mining methods such as, hierarchical clustering or support vector machines are routinely applied in the analysis of high-throughput data coming from DNA microarrays or mass spectrometry [32]. There is a potential of predictive data mining to infer clinically relevant models from molecular data and to therefore provide decision support in the novel field of genomic medicine [32]. Also in a recent research, the authors identify data mining techniques as a potential means of identifying patterns of bias in missing data [33]. Medical data mining can be most rewarding for finding an answer. It can help in extending a life, or giving comfort to an ill person. The rest of the paper is organized as follows, Section 2, presents the background and motivation. Section 3, presents an overview of human-system interactions, Section 4 considers data mining for end-user needs. In Section 5, an integration approach is considered for the above mentioned items. An algorithm is discussed in Section 6. Section 7 considers performance issues. Section 8 presents the discussion and finally, Section 9 describes the summary and conclusions.

2. Background and Motivation

The traditional methods of maintaining health records are being replaced by electronic health records. The EHR revolution is set to facilitate the use of a query language to access and extract knowledge from EHR. It has potential to support additional medical-care related activities such as, decision support and quality management (activities 1, 2 and 3 in **Fig. 3**). Pattern discovery can form an essential support system for clinical decision support systems driven by live data on the actual population [25]. Recent research evaluates the predictive capacity of the clinical EHR of a large mental healthcare provider (75,000 distinct clients a year) to provide decision support information in a real-world clinical setting. Initial research has achieved a 70 percent success rate in predicting treatment outcomes using these methods. More specific activities have also been proposed.

Example (The EU-ADR project): It is funded by the European Commission. It's aim is developing techniques that allow mining of EHRs for adverse drug events across different countries in Europe [24].

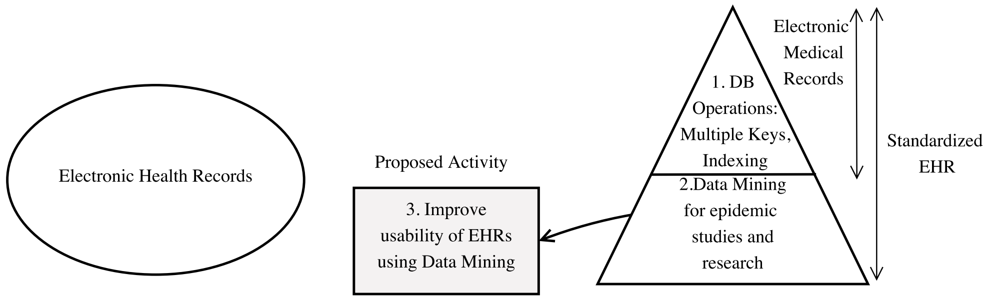
EHR systems must provide support for the data-centric approach. These data oriented systems need to focus on the data quality [*1]. In this study, we focus on the DQ framework for an EHR system, by defining a process for evolving user-friendly EHR systems. With the DQ metrics, various measures can be obtained along with various dimensions for the analysis. We analyze the key aspects of a good EHR system and how these objectives can be met. Earlier efforts in research have concentrated on the secondary use of data mining techniques [2], [9].

The paper proposes a conceptual model that focuses on the steps of a data quality framework. It provides a methodology of how an EHR system can be designed to meet a medical user's needs. A new form of data mining, can be used to mine the user needs to enhance the EHR system usability. Figure 3 shows the evolving potential for the adoption of the Total Data Quality Framework in the case of EHR systems.

3. Human-System Interactions within EHR Systems

Users of EHR systems complain about their difficulties, in terms of the non-friendliness of the systems and about the non-

[*1] Data Quality (DQ) framework is a tool for the assessment of data quality within an organization [21]. It has four components. In a Total Data Quality Management (TDQM) framework [22], the 'Define' component identifies the important data quality dimensions and the corresponding data quality requirements. The 'Measure' component produces the DQ metrics. The 'Analyze' component identifies the root causes for DQ problems and calculates the impact of poor quality information. The 'Improve' component provides techniques for improving DQ.



i) Paper-based Medical Records <----->	Chronological Processing	[Individual patient hospitalization records]
ii) Electronic Medical Records <----->	Chronological search Keyword and index search	[Search restricted to medical records of hospitals] (Activity 1)
iii) Electronic Health Records <----->	Chronological search Keywords and Indexing Data Mining Decision Support	[Patient's lifetime EHRs, epidemic studies on population,public health issues,community health issues] (Activity 1-3)

Fig. 3 Access patterns for an EHR system: activities 1-3.

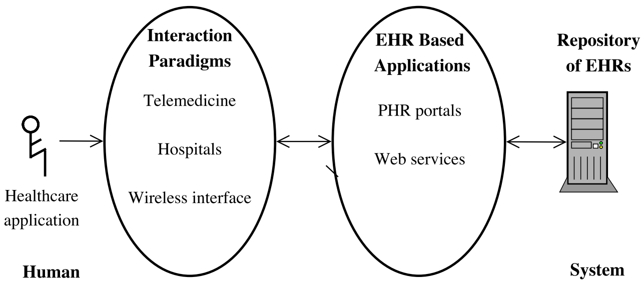


Fig. 4 Human system interactions in EHR domain.

intuitive support for data input [5], [7], [11], [12]. Ideally, the user need not scan or need to search through the entire health record. The user must be able to look at an application layer that accesses and presents a focused amount of information from a given health record or set of records. Unfortunately, the current EHR systems are not optimized for this purpose. Instead, many of them function as something closer to an electronic version of the paper record.

Figure 4 presents the human system interaction in the health care domain with respect to an EHR system. Currently, the healthcare user may interact through various paradigms such as, telemedicine, hospitals and wireless interfaces. The skinput, gaze-based and multitouch technologies may also be used. These interact with the system through various web services as in the case of Personal Health Record (PHR) portals such as Google-Health [12] and Microsoft HealthVault [13].

According to the human-computer interaction principles, the user needs to play a major role in the design of such systems whose ultimate purpose is to be useful for the end-users. Also, the professionals and patients should be treated at the same level. And these interfaces should provide the relevant information and the screens to help them. The user must be able to perform the tasks without having to understand the complete functionality of the system and with a minimal number of clicks to save time [3]. In summary, a user centric EHR system must be created and personalized according to the end-users with the goal to reduce errors and improve learnability, memorability, user satisfaction, and effectiveness of the EHR system.

4. Pattern Discovery: Mining End-User Needs

Considering an EHR system's design at the initial level, the pattern discovery techniques can target the following areas or functions.

(1) Segmenting end-users into groups with a similar age, gender, demographic variables, user groups, purpose, workflows and task-flows;
(2) Understanding the lack of interest among users while using an EHR;
(3) Anticipating end-user's future actions, browsing patterns, reports demanded based on their history and characteristics as per the above segmentation; and
(4) Predicting operations that are needed.

Existing research studies [2], [3], [9] talk about the manual methods such as, observing the end-users on sites and recording their behavior and choices, through questionnaires and surveys as a means to capture data, for end-user involvement in design. Most often, the hidden values and the data mining of resources have the potential to predict trends and the user-behavior.

In the absence of any efforts, these possibilities largely go untapped [8]. The existing literature does not discuss about how the data, that has been collected about the end-user needs (using the above methods), can be incorporated in the development process of the EHR systems. Neither do they give any details on how the intended task flow forecast (or anticipation of browsing patterns exhibited by the users) can be done depending on the different end-user groups. The current study is the first initiative towards these efforts. The main categories of pattern discovery techniques considered for the enhancement of human interaction with an EHR system are described in **Table 1**.

The use of the above techniques has been emphasized in the proposed human system interaction model. **Figure 5**, presents a human system interaction model within the EHR domain. Its aim is to refine the design ideas that test the best options through multiple iterations. It contains a cycle for various steps involved in

Table 1 Pattern discovery for the enhancement of Human Interactions.

	End-User Groups	Loss of interest	Future actions	Ability to learn	Efficiency
Association Analysis	Identify groups and their needs.	Access pattern in logs (extract transactional values).	Users correlation with actions, (for navigation and personalization).	System navigation patterns of a group (facilitates customization).	Map user groups with: functionalities and entities.
Classification and Prediction	Predict a new class using factors such as, functionality, age, sex and education.	Discover complex sections in the EHRs (find tasks: difficult to proceed with).	Find the future action of a user using classification factors.	Capture the user groups on the basis of predefined characteristics.	Improve the functionality based on user groups.
Clustering Analysis	1. Discover classes using the patient data and transaction data. 2. Discover the user choice of operations 3. Discover user-groups.	Discover patterns in user navigation (where the user faces a difficulty).	Find frequent actions (customize according to data entry and retrieval formats).	1. Find patterns of customization, personalization and navigation 2. Use characteristic attributes that were clustered.	Identify end-user groups and the functionality of user groups.
Evolution Analysis	Detect change and growth in information needs.	Analyze the evolving nature of the EHRs for a complexity (unhandled scenarios).	Present the user screens which aid in decision making.	1. Capture the trends in incoming data. 2. Capture the temporal data and find patterns.	Capture the trends within a temporal framework.

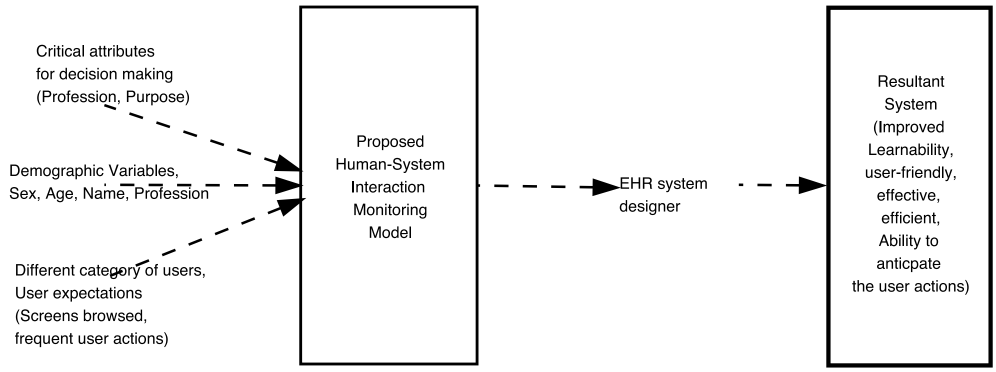


Fig. 6 Proposed Human-System interaction monitoring model for Healthcare Applications.

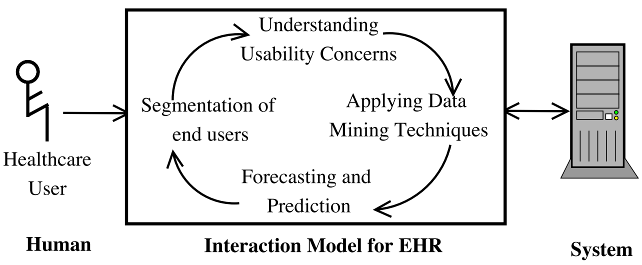


Fig. 5 Human-System interaction monitoring model for EHR systems as per the TDQM framework.

the interaction model. The proposed model can be placed in the domain of healthcare applications as shown in **Fig. 6**. The model will take various user inputs and extract the knowledge that will help the designer to evolve a more efficient EHR system. The performance of the resultant EHR system will be more efficient in terms of time with: a reduced number of clicks, prior knowledge of complex sections of the EHR system, a customized navigation, a user-friendly interface of the EHR system, and a reduced cost of improving the EHR system developed.

There are no differences in the type of system as far as the proposed interaction model is concerned for a different type of medical specialties. The various steps involved in the model (Fig. 5) also remain the same. The choice of pattern discovery techniques often varies depending upon the usage.

5. Improving Human-System Interactions

Usability is defined as the measure of the ease with which a system can be learned and used, including its safety, effectiveness, and efficiency [10]. In order to realize usability improvements on a continuous basis, the user centric design (UCD) approach has been selected. UCD is used for developing applications that incorporate user-centered activities throughout the development process. The TDQM framework can be adopted for improving the quality of user interactions. The TDQM framework has been evolved for development processes, for data and information quality needs [22].

The UCD approach allows end-users to influence the design to increase the ultimate usability of the EHR system [6]. It involves assessing the intended users, analyzing tasks and requirements, testing prototypes, evaluating design alternatives, resolving usability problems, and testing interfaces with users (in an iterative manner). It improves the system's quality because of a more accurate assessment of user requirements and a higher level of user acceptance.

Formerly in 1985, Gould and Lewis [4] introduced three guiding principles of UCD: i) focus on users and tasks early and throughout the design process, ii) measure the usability empirically, and iii) design and test the usability iteratively. Subsequently, standards and techniques for applying these principles have been evolved to meet the needs of specific projects. In the same context, as per the TDQM, the human interaction processes need to be monitored over the entire life cycle.

In order to automate the monitoring of human-computer interactions in the health care domain, we propose a methodology to interpret and discover unknown patterns and relationships between the end-users and functional requirements to improve the usability of EHR systems. It requires the following steps to implement recommendations of a TDQM framework (Fig. 5).

(1) ***Segmentation of end-users***: Segmentation of end-users into different classes is needed to distinguish between the navigation patterns and the expected functionality on the basis of the core characteristics of a group. The users can be segmented into groups on the basis of age (young-old), gender (male-female), demographic variables (education, open to exposure), user groups (patients, insurers, pharmacists, clinicians, nurses), purpose (preventive or diagnostic care), workflows (order in which they may perform a process steps), and task-flows (order in which tasks will be taken by an end-user). For example, the young patients will be more computer-oriented and well-versed with the functionalities even if the system is complex as compared to the elderly patients. **Figure 7** presents a sample decision-tree for the end-user segmentation. Each internal node represents a test on an attribute. Each leaf node represents a class of users (either a specialist doctor or a patient) accessing the EHR system.

(2) ***Understanding the cause of lack of interest among users***: It requires the knowledge of relationships and factors that cause the lack of interest. A continuous analysis of evolving end-user patterns for the workflow is required. This includes factors such as, the complexity of the system and the discovery of a non-user friendly interface.

(3) ***Anticipating future actions***: It is essential to make a map for the current and future actions of the user. It can help facilitate the design of screens relevant to a user. In addition it can lead to a flexible navigation, the personalization and the customization of user interfaces, accessing multiple patients, the delegation of a responsibility among the medical personnel, and enabling data variations and visualizations. This activity enhances the level of user satisfaction.

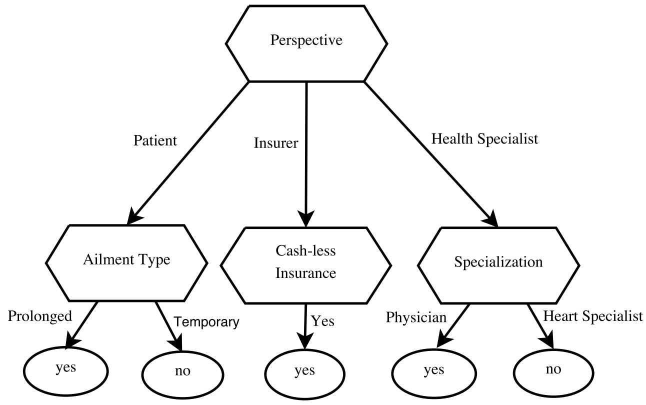


Fig. 7 A decision-tree for end-user segmentation.

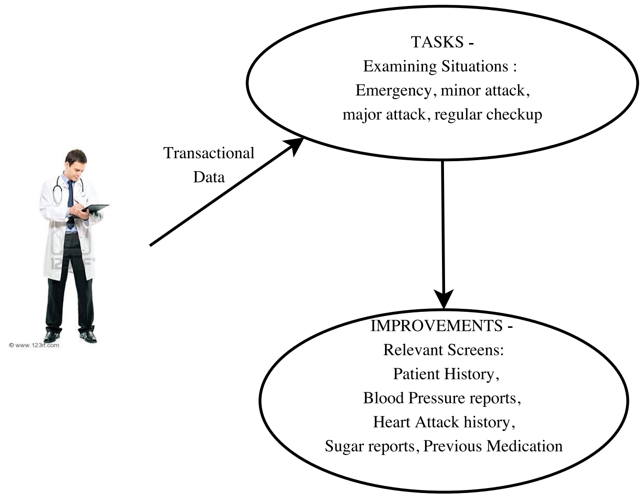


Fig. 8 Monitoring workflow of a specialist's usage of an EHR system.

Table 2 Navigational patterns: Heart Specialist's navigation data.

Transaction ID	Situation	Screens Visited
T100	Emergency	B.P Report, Medication
T200	Regular Checkup	Patient History, B.P Report, Sugar Test Reports, Medication
T300	Major Attack	Patient History, B.P Report, Medication
T400	Minor Attack	Patient History, B.P Report, Sugar Test Reports, Medication

Example 1: Consider the workflow connected with a heart specialist in **Fig. 8**. The relevant data is presented (in transactional form) as shown in **Table 2**. In this study example, an association rule mining algorithm has been applied to find frequent browsing patterns of a heart specialist. Thus, a monitoring system can anticipate the future actions of users and train the system for the user-group specific navigations. The tasks involved and relevant screens shown may vary for a child specialist.

Example 2: Insurers first look for a patient history; next, they look for overall care expenses. Association analysis will enable the EHR system designer to evaluate (i) how often an insurer is expected to follow this flow; (ii) with how much confidence can we state the same.

Such an awareness of browsing preferences in the design phase of the EHR system development enables the EHR system to further present a customized navigation to the end-user.

(4) ***Learnability of the system***: The ease with which the users can accomplish their tasks (the first time they encounter the design) is defined as 'learnability'. For the EHR system, the learnability difficulties need to be identified. This facilitates the design improvements for new users.

(5) ***Efficiency and Effectiveness***: An efficient EHR system must be able to recover from the errors made by the user. Secondly, the system will be able to handle different end-users and provide them for customized navigation and functionalities.

The usability concerns of the EHR system are directly related to the complexity of the system. These require the anticipation of user actions and the ability to learn. We have considered the major categories of pattern discovery techniques (Section 4) for enhancing the EHR system usability. For medical emergencies, wherein the patient is kept under continuous observation, an EHR system trained to record temporal data can be significant for the end-users (such as a heart specialist and a neurologist). Hence, the evolution analysis can be critical in improving the EHR system performance. The above stated categories of techniques, provide support to understand learn ability, and anticipation of future user actions. In the following section we present a procedure for pattern discovery in EHRs.

6. Pattern Discovery

The proposed model aims at refining and extracting the useful information required as inputs to improve EHR system usability. It requires the surveys, questionnaires and the interview data related to user demographics, user expectations and hospital health records preserved in the system. Considering example 1 above, an association analysis algorithm such as the apriori algorithm *2 can be used to find frequent item-sets of the usage patterns of the end-users within an EHR system from the navigationalPatterns database (**Fig. 12**). It contains user actions and navigations within an EHR system in the form of transactions. The model has the capability to find the complex and blocking sections of an EHR system. These can be found by using the navigationalPatterns (Fig. 12) as the training sample and the frequency count of the occurrence of a pattern. A classification or a clustering technique can discover the complex patterns of an EHR system. As an example, the proposed methodology below uses the decision tree induction method for this step of the model.

The resultant knowledge from the above pattern discovery techniques consists of end-user segments, frequent browsing and navigation patterns and identifies the complex sections of the EHR system. These are a part of steps of the conceptual model proposed in (Fig. 5).

This knowledge can be used for forecasting, predicting and learning about user actions for a new class of users coming in, user interface preferences. These act as an input for designing better and efficient EHR systems. The performance of the resultant system will become more efficient in terms of the reduced number of clicks, the prior knowledge of complex sections of the system, the customized navigation and user-friendly interface. The methodology leads to a reduced cost of improving the EHR system being developed. It requires knowledge discovery procedures to operate on a continuous basis. The above procedure is presented in the form of an algorithm: UCD_DM_EHR.

6.1 The Algorithm

Algorithm: UCD_DM_EHR. Find patterns using data mining techniques for improving the usability of an EHR system.

Input:

(1) Training samples (questionnaires, survey reports, existing browsing patterns, hospital archives)

(2) navigationalPatterns (containing usage patterns of the end-users)

(3) Discrete-valued attributes (purpose, region, age, sex)

Output: Consolidated knowledge related to user-groups, user expectations and needs, most frequent browsing patterns within the EHR system and complex, blocking sections of the EHR system.

Method: For the UCD_DM_EHR algorithm, **Fig. 9** defines the procedures for 'create-decision-tree' and 'find-freq-itemsets'.

(1) Perform end-user segmentation using the samples data set and attributes as input
- Create a decision tree to classify the end-users on the basis of the test attributes Fig. 7
- Call procedure **create-decision-tree** (samples, attributes)

(2) Anticipate user concerns and future navigation actions using navigationalPatterns as input
- Find frequent itemsets (screens most browsed, usage patterns) using association analyses
- Call procedure **find-freq-itemsets** (navigationalPatterns)

(3) Predict the complex sections of the EHR system using navigationalPatterns and attributes as input
- Create a decision tree using user actions as test attributes and sections as training data
- Call procedure **create-decision-tree** (navigationalPatterns, attributes)

(4) Forecast and train the EHR system for learn ability and efficiency
- Store results of steps 1, 2, 3 of the above method, for decision making and enhancing EHR system usability.

The above algorithm is given for the most common situations encountered in the process, where a training sample is available, test attributes are known for the classification and all navigational patterns in the form of transactions.

The algorithm UCD_DM_EHR can be enhanced depending upon whether the test attributes are available or not (replace the classification technique with the clustering technique). Whether the transactions are single-dimensional or multi-dimensional, the apriori algorithm can be replaced by an algorithm with the ability to cater to multi-dimensional transactions. Likewise, the complex sections of the system can be discovered using a clustering technique.

*2 Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation) [14], [31].

```
procedure create-decision-tree (training-sample, discrete-attributes)
{
  // If only one discrete-attribute or all samples belong to same class
       // return the discrete attribute as end-user group
  // Else select test-attribute with highest information gain
     // Partition sample data accordingly
  // Repeat the above step, till samples are classified or no further division can be performed
}
procedure find-freq-itemsets (trans-db)
{
  // Scan all transactions to find frequent item-sets of length 1
  // Among the most frequent 1-itemsets find the frequent item-sets of length 2
  // Repeat the process till frequent item-sets of length 'k' are determined
  // At each level the infrequent item-sets are left while moving to the next level
  // The value of 'k' is determined with respect to the number of transactions
}
```

Fig. 9 Procedure for 'create-decision-tree' and 'find-freq-itemsets'.

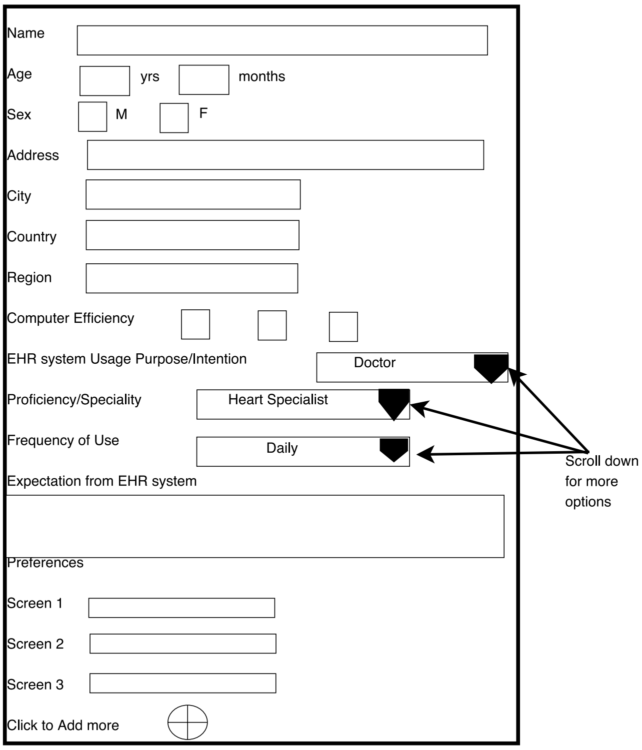


Fig. 10 Sample questionnaire for assessing the usability needs.

6.2 Demonstration

To test the proposed model a representative test case has been prepared with sample data. Weka, an open source data mining tool [29] has been utilized for pattern discovery. The obtained results were evaluated in terms of their usefulness.

6.2.1 Data Preparation

The first step in the Knowledge and Data Discovery process is data preparation [28]. A small data set representative of the real world data collected through questionnaires has been formulated. None of the rows contains any missing values. Since, occurrence of a missing value may generate incorrect association rules. Therefore, care is taken to select data with no missing values. **Figure 10** shows a sample questionnaire. The data has been categorized into two data sets, the userData and the navigationalPatterns as shown in the **Fig. 11** and Fig. 12. The userData is a five-dimensional data set with 24 records containing numerical and categorical attributes. It contains the details of profiles of the potential users of the EHR system, the sex, the age, the region, the purpose and usage class (referred as class). The class attribute is critical and signifies the end-user categories. For our test data, we assume the following classes: nursing, physician, administration, patient and dentist. The second data set navigationalPatterns is a six-dimensional data with 100 records. It contains the categorical attributes. This set contains the browsing preferences of the dentist and physician user classes in various situations. The browsing preferences are marked as 'y' and 'n' for simplification purposes. It represents whether a user wishes to visit a particular screen or not.

6.2.2 Methodology

In this subsection we describe the steps of the algorithm (Section 6.1). The first two steps of the algorithm are critical for knowledge extraction. For the userData data set, the classification has been done with 'class' (end-user category) as the attribute to be predicted using the J48 algorithm [*3] for the classification of records [14]. The minimum support count has been set to 1 (to capture every node formed) for each leaf node. This will enable to capture each user category which accesses the EHR system. The Apriori algorithm has been applied to find frequent rules for the navigation patterns of the user on the second data set. The support to find the rules is kept as 0.55 percent (any rule having more than 50 percent of the occurrences is considered as frequent) and the most frequent itemsets have been discovered as shown in **Table 3**. All the rules were considered, not just the maximal rules are considered.

6.2.3 Results

The result of the categorization of end-users is a decision-tree of size 11 with 6 leaf nodes (**Fig. 13**). The results have 80 percent accuracy and the root mean squared error is 0.2558. Therefore, the first step of the algorithm has exhibited a high accuracy.

From the most frequent itemsets found by the association rule mining of the browsing details of end users, the navigation rules of varied length have been found as shown in the result snapshot **Table 4**. The support statistic used in the apriori algorithm is a sensitive parameter and can be increased or decreased to trace the frequent or infrequent navigational patterns. By decreasing the

[*3] The J48 decision tree classifier, in order to classify a new item first creates a decision tree, on the basis of available training data. It identifies the attribute that clearly classifies the data, its major aim is a maximum information gain. It has an accuracy level and is independent of the volume of data.

```
@relation user_data
@attribute 'Age' { 'OLD', 'YOUNG'}
@attribute 'Is-male' { 'n', 'y'}
@attribute 'Region' { 'ASIA', 'NON-ASIA'}
@attribute 'Purpose' { 'diagnostic', 'preventive'}
@attribute 'Class' { 'administration', 'nursing', 'physician', 'dentist', 'patient'}
@data
'OLD','n','ASIA','diagnostic','dentist'
'YOUNG','y','ASIA','diagnostic','dentist'
'OLD','n','NON-ASIA','preventive','administration'
'YOUNG','n','NON-ASIA','preventive','administration'
'OLD','y','ASIA','diagnostic','dentist'
'OLD','y','ASIA','diagnostic','dentist'
'OLD','y','NON-ASIA','preventive','physician'
'OLD','n','ASIA','preventive','patient'
'YOUNG','n','ASIA','diagnostic','nursing'
'YOUNG','n','ASIA','preventive','patient'
'OLD','y','ASIA','diagnostic','nursing'
'YOUNG','y','ASIA','preventive','patient'
'YOUNG','n','ASIA','preventive','patient'
'OLD','n','NON-ASIA','preventive','administration'
'OLD','n','ASIA','diagnostic','dentist'
'YOUNG','y','ASIA','preventive','patient'
```

Fig. 11 Data set for user details: The userData.

```
@relation sample_transaction
@attribute 'PatientHistory' {'n','y'}
@attribute 'BPReport' {'n','y'}
@attribute 'medication' {'n','y'}
@attribute 'sugar-test-report' {'n','y'}
@attribute 'Class' {'physician','dentist'}
@attribute 'Situation' {'HeartAttack','Cold'}

@data
'n','y','n','y','physician','HeartAttack'
'n','y','n','y','physician','HeartAttack'
'y','y','y','y','physician','Cold'
'n','y','y','n','physician','HeartAttack'
'y','y','y','n','physician','Cold'
'n','y','y','n','physician','HeartAttack'
'n','y','n','y','physician','HeartAttack'
'n','y','n','y','physician','Cold'
'n','y','n','y','dentist','Cold'
'y','y','y','n','physician','HeartAttack'
'n','y','n','y','physician','HeartAttack'
'n','y','n','y','dentist','Cold'
'n','y','y','n','physician','HeartAttack'
```

Fig. 12 Data Set for browsing patterns: The navigationalPatterns.

support below a given threshold the infrequent navigation patterns can be recognized. Analysis of the screens not visited by the end-users based on attributes like language jargon, unfriendly-interface, can point out the complex and redundant sections of the system. Hence, if a user does not exhibit interest in some sections of the EHR system, then the designer can simplify or remove the interface during the development process.

6.3 Analysis of Results

The following examples 3, 4 and 5 present the results.

Example 3: The results in Table 4 show the frequency count of the association rules (the screens or situations in this case). The screens, situation and class are represented interchangeably as antecedent or consequent in the rules. The consideration of these rules can add to the designer's knowledge about which screens are visited by the end-users in a given situation and how often this correlation occurs. Table 4 displays the association of the browsing patterns with a situation for a given user category. For example, as in rule 4, it is 100 percent probable that an end user will check the BP reports in case of a heart attack. Hence, the results shown can be considered by the EHR system developers, as key points to be taken care of while designing the EHR system. This will increase the learnability and predictivity of the system.

Example 4: **Fig. 14** explains graphically a hypothetical usage of an EHR system by an end user (doctor, heart specialist). The left part of the diagram displays an EHR system, where the user has to explore the interface to visit the screens relevant for his purpose (a doctor-patient encounter). Whereas, in the right part, the browsing patterns for the user category are already known and hence, the number of cycles and clicks required to visit relevant screens are reduced. The above situation is a hypothetical situation and shows an example of how the system performance is improved significantly, with the application of the proposed model. When the given model is applied in the real world, the performance will be improved inspite of a rise in the number of screens and the number of users with increase in complexity of EHR systems over their continuous evolution cycles.

Example 5: If a new user registers to use an EHR system; the user will input his purpose, usage criteria for the EHR system along with his demographics. The EHR system has pre-defined categories; it can use the clustering of end-users at the first step of the model and try to include the user in the category which resembles the most. The interface will (at the onset) be customized for that user category. Thus the user will be presented with the relevant screens. For a new user, the complex sections would be removed or simplified. Hence, the EHR system will exhibit the ability to forecast and predict the new user actions. This will result in a simple and smooth user-system experience.

7. Performance Considerations

The interaction model of an EHR is said to learn from pattern discovery (Data Mining (DM)) techniques D with respect to some class of tasks T and performance P, if its performance at tasks in T, as measured by P, improves with DM techniques D. Given training data on the healthcare utilization or the cost, a predictive task (DM technique) may seek to estimate utilization rates

Table 3 Frequent Itemsets discovered from navigational patterns.

Size of Itemset	Discovered Itemsets	Frequency count
4	Patient History=n BPReport=y medication=n sugar-test-report=y	7
4	Patient History=n BPReport=y Class=physician Situation=HeartAttack	7
3	Patient History=n BPReport=y medication=n	7
3	Patient History=n BPReport=y sugar-test-report=y	7
3	Patient History=n BPReport=y Class=physician	7
3	Patient History=n BPReport=y Situation=HeartAttack	7
3	Patient History=n BPReport=y sugar-test-report=y	7
3	Patient History=n Class=physician Situation=HeartAttack	7
3	BPReport=y medication=n sugar-test-report=y	7
3	BPReport=y Class=physician Situation=HeartAttack	8
2	Patient History=n BPReport=y	10
2	Patient History=n medication=n	7
2	Patient History=n sugar-test-report=y	7
2	Patient History=n Class=physician	8
2	Patient History=n Situation=HeartAttack	7
2	BPReport=y medication=n	7
2	BPReport=y sugar-test-report=y	8
2	BPReport=y Class=physician	11
2	BPReport=y Situation=HeartAttack	8
2	medication=n sugar-test-report=y	7
2	Class=physician Situation=HeartAttack	8
1	Patient History=n	10
1	BPReport=y	13
1	medication=n	7
1	sugar-test-report=y	8
1	Class=Physician	11
1	Situation=HeartAttack	8

Table 4 Navigational rules discovered for a physician.

S.No.	Rule	Freq-uency	Confi-dence
1	$Class = physician \Rightarrow BPReport = y$	11	1
2	$PatientHistory = n \Rightarrow BPReport = y$	10	1
3	$sugar - test - report = y \Rightarrow BPReport = y$	8	1
4	$Situation = HeartAttack \Rightarrow BPReport = y$	8	1
5	$Situation = HeartAttack \Rightarrow Class = physician$	8	1
6	$PatientHistory = nClass = physician \Rightarrow BPReport = y$	8	1
7	$Class = physicanSituation = HeartAttack \Rightarrow BPReport = y$	8	1
8	$BPReport = ysituation = HeartAttack \Rightarrow Class = physician$	8	1
9	$Situation = HeartAttack \Rightarrow BPReport = yClass = physician$	8	1
10	$medication = n \Rightarrow PatientHistory = n$	7	1

```
J48 pruned tree
------------------

Region = ASIA
|   Purpose = diagnostic
|   |   Age = OLD: dentist (5.0/1.0)
|   |   Age = YOUNG
|   |   |   Is-male = n: nursing (2.0)
|   |   |   Is-male = y: dentist (1.0)
|   Purpose = preventive: patient (6.0)
Region = NON-ASIA
|   Is-male = n: administration (4.0)
|   Is-male = y: physician (2.0)

Number of Leaves  :     6

Size of the tree :      11
```

Fig. 13 Decision tree for the userData dataset.

for future patients receiving specific treatments.

The DM enriched interfaces provide a new communication-and-control option for individuals for whom the conventional methods are ineffective. There is a performance enhancement using the DM techniques among the various approaches to the EHR design (**Fig. 15**). Arrows in the figure indicate the element that adapts: the EHR, the user or both adapt to optimize and maintain the EHR performance. In the research, the authors proposed a human system interaction Model, in which both the user and the EHR system adapt to optimize and maintain the EHR performance on a continuous basis. The performance can be measured through parameters such as, the sensitivity, the specificity and the predictive accuracy. These parameters have been calculated using the formulas given in Ref. [35]. A high value of sensitivity and specificity imply that the various classes in the userdata dataset are accurately identified with a negligible error. The high predictive accuracy confirms the results. The performance measures are feasible for all the steps but, due to space constraints, we demon-

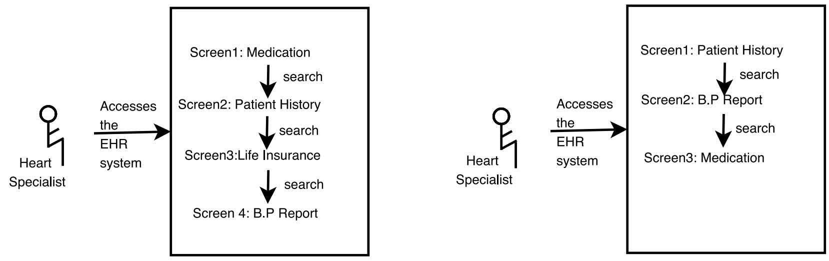


Fig. 14 A graphical view of improved efficiency of the EHR system.

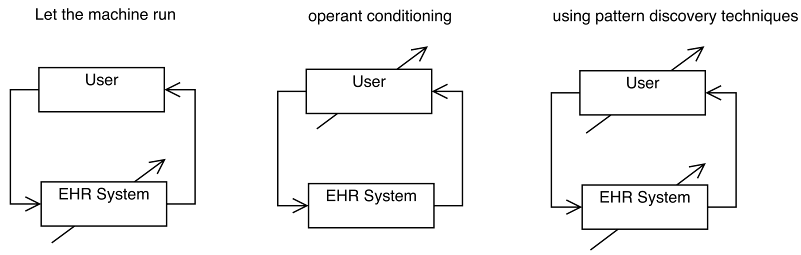


Fig. 15 Approaches to EHR Design.

Table 5 Performance Measures for the classification of end-user categorization.

Class	Administration	Nursing	Physician	Dentist	Patient
Sensitivity	100	75	100	83	86
Specificity	100	95	95	95	95
Predictive Accuracy	100	96	96	96	100

strate the performance for the classification step of the algorithm as given in **Table 5**.

The interaction model will not vary according to the application. The interaction model is based on the strong notions of TDQM (Define, Measure, Analyze and Improve). As per the TDQM, the human interaction processes need to be monitored over the entire life cycle. The various data quality or usability metrics determine the performance of the interaction model which does not depend upon an application. The performance of the healthcare and medical applications (and not the computer systems) will be enhanced on account of the integration and the preservation of the life-time records in a unified way that can be accessed and shared by all physicians and applications, for the entire life time of a patient.

8. Related Studies and Discussions

Medical Systems (and applications) utilize patient records at the hospitals. Similarly, healthcare applications include medical insurance, employee's annual health check records at the company. The integration of medical and healthcare applications is being proposed. Webster's Dictionary defines healthcare as follows [*4] "Healthcare as a general term refers to the delivery of medical services by specialist providers, such as midwives, doctors, nurses, home health aides, vaccination technicians and physician's assistants. Usually such services receive payment from the patient or from the patient's insurance company, although they may be government-financed or delivered by charities or volunteers, particularly in poorer countries". "Social Care" seems to concentrate on the delivery of help to the individual and thus is more indicative of the citizen-centric focus. It includes services such as, fostering Services, family centers, home care, Assessment and Care Management (for older people), Nursing Homes, services for Asylum Seekers and those with HIV/AIDS.

The Microsoft Connected Health Framework (CHF) is designed to enable the creation of Health and Social Care systems that are seamless and joined up. It provides the seamless experience through flexible user interfaces, driving dynamic, orchestrated business processes. It provides the "joined-up" environment by linking applications using open standards for communication, data representation, and process control. Microsoft is involved in projects around the world sharing goals that the Connected Health Framework addresses, including patient safety, improved health, productivity and service delivery reform. Health and Social Care applications, built by independent software vendors and often implemented by system integration partners, can fully leverage the Connected Health Framework to support their legacy applications and the development of future applications [34].

Data interchange between medical, healthcare and welfare domains utilizes industry standards such as HL7, DICOM, and HTTPS [26]. The proposed integration based on the EHRs requires accesses by large variety of professionals and users such as patients.

Designing effective user interfaces is one of the challenges in healthcare. According to the National centre for health statistics, 2009 (mail survey), 43.9 percent of the physicians in USA reported using EHR/EHR systems in their office-based practices. Accurate and timely information access is important for effective EHR systems [30].

This study examines UCD and TDQM recommendations, re-

*4 http://www.websters-online-dictionary.org/definition/healthcare lated work regarding the usability of EHR systems and data mining techniques. It proposes to integrate these aspects to increase the benefits for the healthcare users. Thus, the paper first targets the key areas of design where the user can be involved in the development cycle of the EHR systems. Next, it analyses the scope of automation for the proposed model to overcome the challenges.

The model will be a pre-process to the design of the EHR system. The use of the algorithm will reduce the cost, the time and the effort in the development and the utilization of the EHR system. It will maintain the efficiency and the effectiveness of the system in view of the increase in complexity in the medical domain on a continuous basis. The gap between the user requirements and the designer's vision is reduced with the application of this model. It is adopted for the entire healthcare application or the EHR system.

9. Summary and Conclusions

The standardized EHR proposals consider a maximal content in order to include the requirements of differences in type of system (medical specialty, etc.). For example, the Archetype for heart rate Fig. 2. The EHRs facilitate the archiving of data over the lifetime of patients. Although, depending upon the requirements, each medical application adopts different processes and procedures. Data mining techniques can provide useful insights for complex queries raised by medical researchers.

The proposed technique uses EHR contents and access logs for pattern discovery. It does not add overheads on day-to-day working. The cost introduced by an inefficiency within a complex system, must be eliminated. Currently, in the absence of TDQM based quality improvements, these cost and difficulty become a critical factor, with a growing size and complexity. Our objective is to minimize the expected cost of user's interactions or (equivalently) to maximize the user's expected utility. The decision-theoretic optimization provides a flexible and principled approach for these systems. The quality of the resulting solution is dependent on the accuracy of the underlying utility or cost function. Unfortunately, determining the correct utility function is a complex, time-consuming, and error-prone task. The data mining techniques facilitate to parameterize the UCD_DM_EHR algorithm and then gain the DQ perspective in terms of accuracy, ease and timely response. Thus, the healthcare workers are relieved of the laborious and unreliable process of hand-tuning.

There is a large scope to explore the specific algorithms or approaches that can be applied to achieve an accurate capturing of the desired functionality. Using this framework, we can automate the process of capturing human needs for an end-user system in the healthcare domain. Further, in a user centric interaction model, there is no need to find features that users do not need.

The performance of a computer is less significant, in view of the proposed system integration. The performance may be viewed with respect to the accuracy, the efficiency and saving of time and cost with respect to the integration of medical and healthcare applications. Also, the overall integration will reduce the cost due to larger scale operations and the removal of duplicated efforts.

Once the EHR system designer incorporates the application of data mining techniques, it will enable the learnability of the system. The major advantage of this model is the scalability. The algorithm is scalable with a minimal error rate, so large data sets can be mined.

As a conclusion, an Electronic Health Record involves all the information and procedures that healthcare professionals perform in treating a patient. These offer sufficient data for decision-making in both the clinical service of a patient care and the administration of public health care. This is a situation similar to the formation of a Dataware house (DWH) with patients' data along with time and space dimensions. Thus, the development of EHR systems helps to enhance the quality of patient care and the efficiency of clinical services.

The evolution of EHRs can promote the ease of use of the health care systems. At the same time, the internal complexities of the system have increased. As the amount of data becomes voluminous, there is an increase in the need for interoperability, epidemic studies and to provide a user-friendly system. The traditional approach should be replaced for EHR systems by involving the principles of a TDQM framework [22]. In the healthcare domain this is essential to reduce health risks with each change and within the process of evolution. For this purpose, understanding of user needs, their inhibitions to use the system, and predicting their navigation patterns are required. The traditional methods such as, observing the user, filling up surveys can aid this requirement. These efforts are not efficient and have a limited scope. Thus, the proposed framework targets the design phase of an EHR system.

This study proposes the use of various techniques to find the unknown usage patterns. It determines their correlation to improve the usability of the EHR system. Mapping representative classes of users to their respective tasks has been emphasized. Thus, a human-system interaction model has been identified for the EHR domain (Fig. 5). Further, research on improving the user satisfaction by performing a medical interaction effectively and efficiently will be required. The proposal can be applied to a complex system or domain to increase the productivity in terms of interaction. These efforts will lead the developments in using similar techniques to build intelligent software for the EHR systems and the related bio-medical domain.

Reference

[1] Smelce, J.B., Miller-Jacobs, H. and Kantrovich, L.: Usability of Electronic Medical Records, *Journal of Usability Studies*, Vol.4, No.2, pp.70–84 (2009).

[2] Ramakrishnan, N., Hanauer, D. and Keller, B.: Mining Electronic Health Records, *Computer* (*IEEE Magazine*), Vol.43, No.10, pp.77–81 (2010).

[3] Dabbs, A.D.V., Myers, B.A., Mc Curry, K.R., Dunbar-Jacob, J., Hawkins, R.P., Begey, A. and Dew, M.A.: User-Centered Design and Interactive Health Technologies for Patients, *CIN: Computers, Informatics, Nursing*, Vol.27, No.3, pp.175–183 (2009).

[4] Gould, J.D. and Lewis, C.: Designing for Usability: Key Principles and What Designers Think, *Comm. ACM*, Vol.2, No.3, pp.300–311 (1985).

[5] Abras, C., Maloney-Krichmar, D. and Preece, J.: *User-Centered Design*, Bainbridge, W. (Ed.), *Encyclopedia of Human-Computer Interaction*, Thousand Oaks (2004).

[6] Lee, S.H.: Usability Testing for Developing Effective Interactive Multimedia Software: Concepts, Dimensions and Procedures, *Educa-*

tional Technology & Society, Vol.2, No.2, pp.1436–1440 (1999).
[7] Beyer, H. and Holtzblatt, K.: *Contextual Design: Defining Customer-Centered Systems*, Academic Pressm, San Diego (1998).
[8] Rogers, G. and Joyner, E.: Mining your data for Health Care Quality Improvement, *SAS User Group International Conference*, pp.641–647 (1997).
[9] Lin, J.-H. and Haug, P.J.: Data Preparation Framework for Preprocessing Clinical Data in Data Mining, *Proc. 2006 AMIA Annu Symp*, pp.489–493 (2006).
[10] Mchome, S., Sachdeva, S. and Bhalla, S.: A brief Survey: Usability in Healthcare, *International Conference on Electronics and Information Engineering* (*ICEIE*), Vol.1, pp.463–467 (2010).
[11] Kjeldskov, J., Skov, M.B. and Stage, J.: A longitudinal study of usability in health care: Does time heal?, *International Journal of Medical Informatics*, Vol.79, No.6, pp.e135–e143 (2010).
[12] GoogleHealth, available from ⟨http://www.google.com/intl/en-US/health/about/index.html⟩.
[13] Microsoft Health Vault, available from ⟨http://www.healthvault.com/⟩.
[14] Han, J. and Kamber, M.: *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers (2001), available from ⟨www.cs.sfu.ca/˜han/DM_Book.html⟩.
[15] Agrawal, R., Imielinksi, T. and Swami, A.: Mining association rules between sets of items in large database, *The ACM SIGMOD Conference*, pp.207–216, Washington DC, USA (1993).
[16] Hristidis, V.: *Chapter 3 on Overview of Information Discovery Techniques on EHRs and chapter 7 on Data Mining and Knowledge Discovery on EHRs*, Information Discovery on Electronic Health Records, Chapman and Hall/CRC Data Mining and Knowledge Discovery Series (2009).
[17] Health Level Seven (HL7), available from ⟨www.hl7.org⟩.
[18] Hussein, R., Engelmann, U., Schroeter, A. and Meinzer, H.P.: DICOM structured reporting: Part 1. Overview and characteristics, *RadioGraphics*, Vol.24, No.3, pp.891–896 (2004).
[19] Integrating the Healthcare Enterprise, available from ⟨http://www.ihe.net/⟩.
[20] Eichelberg, M., Aden, T., Riesmeier, J., Dogac, A. and Laleci, G.B.: A survey and analysis of Electronic Healthcare Record standards, *ACM Comput. Surv.*, Vol.37, No.4, pp.277–315 (2005).
[21] Wang, R.Y. and Strong, D.M.: Beyond accuracy: What data quality means to data consumers, *Journal of Management Information Systems*, Vol.12, No.4, pp.5–33 (1996).
[22] Wang, R.Y., Zaid, M. and Lee, Y.W.: *Book on Data Quality*, Kluwer Academic publishesrs (2001).
[23] Cios, K.J. and Moore, G.W.: Uniqueness of medical data mining, *Artificial Intelligence in Medicine Journal*, Vol.26, No.1-2, pp.1–24 (2002).
[24] Trifiro, G. et al.: Data mining on electronic health record databases for signal detection in pharmacovigilance: Which events to monitor, *Pharmacoepidemiology and Drug Safety*, Vol.18, No.12, pp.1176–1184 (2009).
[25] Bennett, C. and Doub, T.W.: Data Mining and Electronic Health Records: Selecting Optimal Clinical Treatments in Practice, *Proc. 2010 International Conference on Data Mining* (*DMIN 2010*), Las Vegas, Nevada, USA, CSREA Press, ISBN 1-60132-138-4 (2010).
[26] Palta, J.R., Frouhar, V.A. and Zlotecki, R.A.: Leveraging Pervasive Technologies to improve collection of prostate cancer outcome data, *Computer*, Vol.43, No.7, pp.35–42 (2010).
[27] Tsymbal, A. and Bolshakova, N.: Guest Editorial Introduction to the Special Section on Mining Biomedical Data, *IEEE Trans. on Information Technology in Biomedicine*, Vol.10, No.3, pp.425–428 (July 2006).
[28] Fayyad, U., Piatetsky-Shapiro, G. and Smith, P.: The KDD process for extracting useful knowledge from volumes of data, *Comm. ACM*, Vol.39, pp.27–34 (1996).
[29] Weka Tool, available from ⟨http://www.cs.waikato.ac.nz/ml/weka/⟩.
[30] available from ⟨http://www.coresolutionsinc.com/UsabilityWhitepaper.pdf⟩.
[31] Welch, S.R. and Huff, S.M.: *Cohort Amplification: An Associative Classification Framework for Identification of Disease Cohorts in the Electronic Health Record*, AMIA (2010).
[32] Bellazzi, R. and Zupanb, B.: Predictive data mining in clinical medicine: Current issues and guidelines, *International Journal of Medical Informatics*, Vol.77, No.2, pp.81–97 (2008).
[33] Tremblay, M.C., Dutta, K. and Vandermeer, D.: Using Data Mining Techniques to Discover Bias Patterns in Missing Data, *Journal of Data and Information Quality* (*JDIQ*), Vol.2, No.1 (July 2010).
[34] Microsoft Health framework, available from ⟨http://www.microsoft.com/industry/healthcare/technology/HealthFramework.mspx⟩.
[35] Cios, K.J., Pedrycz, W. and Swiniarski, R.: *Data Mining Methods for Knowledge Discovery*, Kluwer, ISBN 0-7923-8252 (1998).

Shelly Sachdeva received her B.E. (Hons) and M.Tech. (Hons) degrees in Computer Science, in India in 2001 and 2004, respectively. She has teaching experience of 8 years. She is currently pursuing a Ph.D. from University of Aizu, Japan. Her main research interests are in the area of Electronic Health Record Databases, the high-level Query Interfaces, and Data Quality for Health Informatics.

Aastha Madaan received a B.Sc. (Hons) in Computer Science from Delhi University in 2006. She received a M.Sc. in Computer Science from Delhi University in 2008. She has worked as a software developer with Aricent Techologies Group for almost 3 years. She is currently pursuing a Ph.D. in Computer Science from University of Aizu, Japan.

Subhash Bhalla received his B.Tech. in Computer Science in 1978. He received his Ph.D. degree in Computer Science in 1984 from Indian Institute of Technology, Delhi, India. His research interests include the design of new databases to support multimedia information systems, data modeling and transactions and distributed algorithms. He is currently working as professor at University of Aizu, Japan. He has 24 years of experience as faculty, scientist and researcher.