

Received 3 June 2013; revised 29 July 2013; accepted 15 August 2013. Date of publication 21 November 2013;
date of current version 23 December 2013.

Digital Object Identifier 10.1109/JTEHM.2013.2290113

icuARM – An ICU Clinical Decision Support System Using Association Rule Mining

CHIH-WEN CHENG¹ (Member, IEEE), NIKHIL CHANANI², JANANI VENUGOPALAN³ (Member, IEEE), KEVIN MAHER², AND MAY DONGMEI WANG^{1,3,4,5,6} (Senior Member, IEEE)

¹School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

²Pediatrics Department, Emory University, Atlanta, GA 30332, USA

³The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, USA

⁴Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, GA 30332

⁵Institute for People and Technology, Georgia Institute of Technology, Atlanta, GA 30332

⁶Department of Hematology and Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322

This work was supported by grants from Georgia Tech – Children’s Healthcare of Atlanta, Georgia Cancer Coalition Award to Prof. M. D. Wang, Hewlett Packard, and Microsoft Research.

ABSTRACT The rapid development of biomedical monitoring technologies has enabled modern intensive care units (ICUs) to gather vast amounts of multimodal measurement data about their patients. However, processing large volumes of complex data in real-time has become a big challenge. Together with ICU physicians, we have designed and developed an ICU clinical decision support system icuARM based on association rule mining (ARM), and a publicly available research database MIMIC-II (Multi-parameter Intelligent Monitoring in Intensive Care II) that contains more than 40,000 ICU records for 30,000+ patients. icuARM is constructed with multiple association rules and an easy-to-use graphical user interface (GUI) for care providers to perform real-time data and information mining in the ICU setting. To validate icuARM, we have investigated the associations between patients’ conditions such as comorbidities, demographics, and medications and their ICU outcomes such as ICU length of stay. Coagulopathy surfaced as the most dangerous co-morbidity that leads to the highest possibility (54.1%) of prolonged ICU stay. In addition, women who are older than 50 years have the highest possibility (38.8%) of prolonged ICU stay. For clinical conditions treatable with multiple drugs, icuARM suggests that medication choice can be optimized based on patient-specific characteristics. Overall, icuARM can provide valuable insights for ICU physicians to tailor a patient’s treatment based on his or her clinical status in real time.

INDEX TERMS Intensive care units (ICUs), personalized clinical decision support system, association rule mining, clinical risk prediction models.

I. INTRODUCTION

According to the Society of Critical Care Medicine (SCCM), there are approximately five million patients admitted annually to intensive care units (ICUs) in the United States, with average mortality rates ranging from 10% to 29% [1], which are the highest rates of all the units in a hospital. In addition, the ICU has some of the highest rates of medical errors as compared to most clinical settings due to the complexity of care [2, 3]. With the extensive hemodynamic monitoring and use of multiple measurement technologies, the modern ICU generates large volumes of complex and multimodal data. Interpreting and utilizing this information is challenging for the ICU physician. The data is richer, but the ability to integrate it for effective actions remains difficult.

We have designed and developed an ICU clinical decision support system (CDSS) to improve outcomes in critically ill patients by providing real-time decision support, decreasing medical errors, and minimizing life-threatening events caused by delayed or uninformed medical decisions. CDSSs are computer-aided “active knowledge systems which use two or more items of patient data to generate case-specific advice” [4]. Evidence has strongly suggested that CDSSs can improve a physician’s decision making performance [4]. For optimal medical decision making, the CDSS needs to be data-driven, rapid, and informed.

Evidence-based medicine is the “conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients” [5]. A CDSS is

evidence-based if its knowledge base is derived from, and continually reflects, the most up-to-date evidence from the scientific literature and practice-based sources [6]. A generic form of evidence in an evidence-based CDSS is the IF-THEN rule. The rule implies that IF an antecedent (i.e., a set of conditions) presents, THEN an outcome is expected or an action should be taken. In the literature, evidence-based CDSSs have provided important risk assessment scores for clinicians, such as prediction of ICU survivability (e.g., APACHE II [7]), length of stay [8], organ failure (e.g., SOFA [9]), neurologic prognosis (e.g., Glasgow Coma Score [10]), and outcomes after acute coronary syndrome (e.g., GRACE ACS model [11]). However, none of the above applications are true CDSSs because they are not interactive nor flexible, which are two key features of true decision support systems [12].

These computer-based ICU assistance systems claim to be CDSSs because they provide the “statistics” of evidence. However, it is difficult for clinicians to make a correct decision by recalling all corresponding knowledge based on these statistics in a timely fashion. They have to search their archives, find the appropriate literature, and interpret relevant evidence (assuming it is up-to-date). This decision support process is not feasible in the critical care setting, when the luxury of time is rare. Thus, a reliable CDSS in an ICU should provide not only statistically significant knowledge, but also an interactive user interface that enables clinicians to search for evidence effectively and in real-time.

Besides being interactive, an evidence-based ICU CDSS need to be flexible. In typical ICU systems that claim to be CDSSs, researchers define expected IF-THEN rules (i.e., with conditions and outcomes or actions) given certain clinical problems, validate the rules (i.e., via human trial, lab experiment, or computer simulation), and form new evidence if the rule is statistically significant (e.g., $p < 0.05$). Afterwards, clinicians can refer to the evidence if the clinical conditions are matched. For example, a clinician can select an appropriate antimicrobial drug for a septic patient when the pathogenic organism has specific hemodynamic and biochemical markers. However, some patients lack clear-cut evidence for the presence of an infection and/or the type of infecting organism, which makes the decision to treat with an antimicrobial drug experience-based instead of evidence-based. The clinician still needs to make the same decision about antimicrobial drug prescription with incomplete information, and then passively assess the prognosis. Such a process introduces human bias that deviates from the original design of the CDSS. Clinicians face this challenge on a daily basis for every patient in the ICU due to heterogeneous conditions. Therefore, a flexible ICU CDSS is needed to allow clinicians to customize conditions to better describe a patient’s immediate status (i.e., personalized), instead of referring to fixed evidence formed from different clinical situations.

In addition to the two key features mentioned above, a powerful CDSS relies on a sufficient and representative database

of patient ICU stays. Although the bedside monitor can generate large amounts of data from each patient as compared to other care settings [13], the number of unique patients in a standard ICU CDSS database is typically small. Thus, the accuracy of new decision support evidence is limited by the diversity of phenotypes contained in a small number of patients. Even though some studies contain large numbers of patients, the evidence is usually mined from the entire cohort without clinical categorization. Ideally, before applying data modeling analysis, a CDSS should first extract a cohort of patients who have similar medical histories and situations, and reveal the sample size as a reference when delivering new evidence back to clinicians. Clinicians can judge when the evidence is mined from a sufficient and representative dataset.

To our knowledge, to date, no true CDSSs have been developed for the ICU. To address the aforementioned challenges, we designed and developed an ICU CDSS called icuARM. icuARM has an interactive graphical user interface (GUI) that allows clinicians to mine decision support evidence in a flexible way. We constructed association rules to perform evidence-based data mining from a database with more than 40,000 ICU stays of 30,000+ patients. We will report icuARM as the following structure. First, we provide a short description of data source in Section II.A. Then we describe the detailed data mining approach and the system user interface in Sections II.B and II.C respectively. Next we present results, evaluation and discussion of two case studies in Section III. Finally we summarize the conclusion and future directions in Section IV.

II. METHODS AND PROCEDURES

A. DATA SOURCE - THE MIMIC-II DATABASE

The data in the icuARM is imported from the Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database. MIMIC-II is a publicly accessible ICU data repository containing records of over 40,000 ICU stays in which 32,000 are adult (>15 yrs) records and 8,000 are neonatal (<2 yrs) records [14]. The data in MIMIC-II can be categorized into two major categories: clinical data and physiological data. The clinical data is collected from MIMIC-II’s ICU information systems and hospital electronic health record systems. The high-resolution physiological data consists of time series waveforms and time series measurements from bedside monitors. The data mining process in this study only includes clinical data. In the near future, we will utilize temporal data mining on time series physiological data.

The imported clinical data consists of approximately 232 million entries covering over 13,000 variables. We further divide the MIMIC-II clinical data into two groups of categories: basic and event-based. The basic categories include data that remain unchanged during one ICU stay (e.g., patient demographics and pre-existing comorbidities). The event-based categories contain data collected at multiple

TABLE 1. Description of MIMIC-II clinical data.

Category (mapping)	Event	Measure examples	No. of available variables
Basic (1-to-1)	ICU general data	Number of ICU stays, stay sequence, first/last ICU day flag, stay in/out date, stay length (min), stay death flag, care unit of the first/last ICU day, SAPS ¹ score, SOFA ² .	42
	Demographics	Gender, age, ethnicity, religion, date of birth/death, death flag, marital status, weight, height, admission type.	14
	Comorbidities	Congestive heart failure, hypertension, cardiac arrhythmias, pulmonary circulation, peripheral vascular, paralysis, neurological disorder, chronic pulmonary, cancer, AIDS, diabetes.	32
Events (1-to-N)	³ Medications	Propofol, Insulin, Fentanyl, Heparin, Neosynephrine-k, Levophed-k, Midazolam, Furosemide, Lorazepam, Labetolol, Eptifibatide.	406
	³ Fluid balances	Urine out, D10W, promote with fiber, replete with fiber, lactated Ringer’s, free water bolus, gastric, nasogastric.	6,809
	Laboratory tests	Hematocrit of blood, potassium in serum or plasma, creatinine in serum or plasma, urea nitrogen in serum or plasma, hemoglobin in blood, pH of blood.	714
	³ Nurse-verified charting	Heart rate, heart rhythm, blood pressure, noninvasive blood pressure, central venous pressure, SaO ₂ , arterial PH, arterial PaCO ₂ , arterial PaO ₂ , arterial CO ₂ , SpO ₂ , respiratory rate.	4,781

¹Simplified Acute Physiology Score

²Sequential Organ Failure Assessment

³Event-based categories that include event duration

time points within an ICU stay, including laboratory tests (e.g., blood chemistries, complete blood counts), medication events (e.g., insulin, heparin), fluid balance (e.g., urine output), and nurse-verified chart measurements (e.g., blood pressure, heart rate). Values in event-based categories are processed to generate mean, minimum, maximum, and standard deviation during an ICU stay. Durations of chart measurement, medication, and fluid balance events are also imported. We tabularized the imported MIMIC-II clinical data in Table 1.

B. PRINCIPLE OF ASSOCIATION RULE MINING

After importing data from MIMIC-II, a sophisticated mining process is required to unearth meaningful associations from such a voluminous dataset. Association rule mining (ARM) is a method to reveal meaningful relations between variables in databases. Agrawal *et al.* first introduced the concept of ARM to extract regularities between products in large-scale warehouse databases [15]. In healthcare, ARM has been widely adopted in applications such as heart disease prediction [16, 17], healthcare auditing [18, 19], and neurological diagnosis [20, 21]. However, to our understanding, there is no published CDSS reporting the use of ARM in the ICU.

Rules in ARM are in the form of $X \Rightarrow Y$, which means that X implies Y , where X and Y are called antecedent and consequent, respectively. In its original marketing analysis context, the rule $X \Rightarrow Y$ carries the meaning that if a customer buys items in X , he/she is also likely to buy items in Y . In MIMIC-II, one patient may be associated with one or more hospital stays; and one hospital stay may be associated with one or more ICU stays. Therefore, the most basic data piece in our mining process is the ICU stay. Therefore, in our ICU data mining process, a rule $X \Rightarrow Y$ implies that if X occurs in one ICU stay, Y is also likely to occur during the

stay. Here the ICU stays are analogous to transactions in the traditional marketing ARM in which a customer may have multiple market visits, and a market visit may have multiple transactions.

The antecedent X and consequent Y are itemsets that consist of one or more item(s). An item is composed of a variable with a corresponding value or a range of values. An item can be numerical or categorical depending on the data type of the variable. For example, $HeartRateMax > 80$ is a numerical item that has a variable $HeartRateMax$ and a range of values “ > 80 .” As another example, $IsFirstDay = Yes$ is a categorical item that has a variable $IsFirstDay$ and a value Yes . Because X and Y can consist of one or a combination of items, the associations are not necessarily one-to-one.

Two important metrics — *support* and *confidence* — quantify the frequency and level of association of a rule. We modified these two metrics from their conventional forms to fit our ICU clinical mining [22]. First, the support of an association rule is defined as:

$$Supp(X \Rightarrow Y) = \frac{count(X \cup Y)}{count((X_V \neq \phi) \cup (Y_V \neq \phi))} \quad (1)$$

where $X \cup Y$ indicates the set of all ICU stays in which both X and Y occur; and $X_V \neq \phi$ and $Y_V \neq \phi$ indicate the set of all ICU stays in which all variables in X and Y have no missing values. For example, if $X = “HeartRateMax > 80,”$ $X_V \neq \phi$ refers to all ICU stays in which $HeartRateMax$ has been assigned a value (i.e., no missing data). $count(a)$ returns the number of ICU stays that contain a , where a can consist of one or more items. The numerator of (1) counts the total number of ICU stays that contain all items of $X \cup Y$. The denominator of (1) counts the total number of stays that have no missing data in all variables of X and Y . This is critical in ICU data mining because clinical data is usually

recorded when the patient is presenting a specific condition or undergoing a specific treatment. Focusing on the ICU stays that do not have any missing data in any of the variables of X and Y helps extract stays that are under similar clinical conditions. Therefore, the support of a rule $X \Rightarrow Y$ indicates the fraction of the ICU stays that hold $X \cup Y$ to those that have no missing data in all of the variables of X and Y . The support ranges from 0 to 1. A high support for an association rule indicates that a high portion of ICU stays are applicable to the rule.

Another metric of an association rule is its confidence:

$$Conf(X \Rightarrow Y) = \frac{count(X \cup Y)}{count(X \cup (Y_V \neq \phi))} \quad (2)$$

In all ICU stays that have no missing value in variables of X and Y , the confidence calculates the ratio (ranged from 0 to 1) of ICU stays that match all items in both X and Y to the records that match all item values in X no matter the value of Y . For example, if the confidence of an association rule $\{HeartRateMax > 190\} \Rightarrow \{Death = YES\}$ is 90%, it implies that for ICU stays that have $\{HeartRateMax > 190\}$, 90% of these stays have $\{Death = YES\}$. In other words, confidence reveals the level of the association between X and Y .

In order to discover frequent and confident association rules, the mining process requires users to specify two minimum values as thresholds to drop infrequent and unconfident rules, which are minimum support ($Supp_{min}$) and minimum confidence ($Conf_{min}$). Rules are considered to be frequent if their supports are at least $Supp_{min}$ and confident if their confidences are at least $Conf_{min}$. The goal of ARM is to find all frequent and confident rules based on these two user-specified values.

There are two main steps in revealing association rules. The first step is to find all frequent itemsets that have supports above $Supp_{min}$. The second step is to use the frequent itemsets to generate confident rules with confidences above the $Conf_{min}$. Because the second step is straightforward, most of the research focus is on the first step. Since the first algorithm was introduced in the original report of ARM [15], new algorithms have been proposed to improve the efficiency of the generation of frequent itemsets. Among these algorithms, the Apriori algorithm is the most popular in ARM research.

The Apriori algorithm utilizes an iterative process to generate frequent itemsets. Let $I = \{I_1, I_2, \dots, I_N\}$ consist of N possible items in the database. In the first iteration, the algorithm starts by counting the occurrence of 1-itemset candidates that contain only one item. 1-itemset candidates that have supports lower than $Supp_{min}$ are pruned out and the remaining ones are called frequent 1-itemsets. In the following iterations (i.e., $k > 1$), the candidate k -itemsets are first generated by joining the frequent $(k-1)$ -itemsets. Then frequent k -itemsets are generated by pruning out candidate k -itemsets that have supports lower than $Supp_{min}$. The iteration continues until no more candidates or frequent itemsets can

be found. The pseudo-code of the Apriori algorithm presented in [22] is given as follows:

Algorithm: $F = Apriori(T, I, Supp_{min})$
// **Input:** T (Transactions), I (1-itemsets), $Supp_{min}$
// **Output:** F (Frequent Itemsets)
 $F_1 = \{f \mid f \in I, f.support \geq Supp_{min}\};$
for $(k = 2; F_{k-1} \neq \emptyset; k++)$ **do**
 $C_k = GenCandidate(F_{k-1});$
 for each transaction $t \in T$ **do**
 for each candidate $c \in C_k$ **do**
 if c is contained in t **then**
 $c.count++;$
 end
 end
 $F_k = \{c \in C_k \mid c.support \geq Supp_{min}\}$
 end
return $F = \bigcup_k F_k;$

The *GenCandidate* in the Apriori algorithm is the candidate itemset generation algorithm that is given as follows:

Algorithm: $C_k = GenCandidate(F_{k-1})$
// **Input:** F_{k-1} (Frequent $k-1$ itemsets)
// **Output:** C_k (Candidate k itemsets)
 $C_k = \emptyset;$
forall $f_m, f_n \in F_{k-1}$
 where $f_m = \{i_1, \dots, i_{k-2}, i_{k-1}\}$
 and $f_n = \{i_1, \dots, i_{k-2}, i'_{k-1}\}$
 and $i_{k-1} \neq i'_{k-1}$ **do**
 $c = \{i_1, \dots, i_{k-1}, i'_{k-1}\};$
 $C_k = C_k \cup \{c\};$
 foreach $(k-1)$ -subset s of c **do**
 if $(s \notin F_{k-1})$ **then**
 delete c from $C_k;$
 end
end
return $C_k;$

After generating all frequent itemsets via the Apriori algorithm, the second subproblem is to generate confident rules that satisfy $Conf_{min}$. For each frequent itemset f , consider all non-empty subsets of f . For each subset a , the process forms a new rule $a \Rightarrow (f - a)$ if its confidence is above $Conf_{min}$. We can then call a and $(f - a)$ the antecedent and consequent, which are the X and Y , respectively, of an association rule.

As described previously in (2), we can assess if a rule $X \Rightarrow Y$ has a high level of association according to its confidence. However, a high confidence of rule $X \Rightarrow Y$ still cannot guarantee a low confidence of its counter case. That is, during clinical decision making, we want to consider when the rule $X \Rightarrow Y$ yields a higher confidence than its counter case. This means that Y is likely to occur only when X occurs, and when X does not occur, Y has a low chance of occurrence. We use the following equation to evaluate the importance of

a rule $X \Rightarrow Y$:

$$Impo(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{conf(\bar{X} \Rightarrow Y)} \quad (3)$$

The importance metric ranges from 0 to ∞ . A rule of importance less than 1 means that the antecedent predicts the consequent worse than the counter case of the antecedent. This type of rule should be ignored. Thus the rules are expected to have an importance >1 . To avoid the rules with importance close to 1 (i.e., 0.9 and 1.1) due to random chance, a more strict and higher threshold for importance are used to ensure statistical significance. In this study, the qualitatively chosen threshold of importance is selected to be >1 .

We can utilize the aforementioned three metrics to determine if a rule $X \Rightarrow Y$ is frequent, confident, and important. However, a given consequent Y may be associated with different antecedents from different rules. For example, if X_1 and X_2 are both possible antecedents that associate with a consequent Y , we can emphasize X_1 if the presence of Y is dominated by X_1 . Therefore, we use a new metric to determine the dominance of an antecedent on a rule's consequent.

$$Domi(X \Rightarrow Y) = \frac{count(Y \cup X)}{count(Y \cup (X_V \neq \phi))} \quad (4)$$

Similar to (2), the dominance ranges from 0 to 1. The dominance rule of $X \Rightarrow Y$ is identical to the confidence of $Y \Rightarrow X$ because the rule can be viewed as the ratio of ICU stays that match all items in X and Y to the records that match all items in Y no matter the value of X .

Because the antecedent of a rule can consist of multiple items, we may also need to determine how a new item affects a rule's confidence when we include the item in the rule's original antecedent (i.e., $X \cup I_{New} \Rightarrow Y$). Evaluating effects of new clinical items would be helpful, for example, in making decisions about medication or treatment combinations. Additionally, it is not necessary to always pursue positive or negative effects; rather, it depends on the clinical situation. For example, when the consequent of a rule is the mortality of an ICU stay given the antecedent of a medication X , we may not only be interested in drugs that may decrease mortality, but also in those that may increase mortality. In this study, we use the following equation to measure the effect of adding a new antecedent item on a rule's confidence:

$$Effe(I_{New}|X \Rightarrow Y) = conf(X \cup I_{New} \Rightarrow Y) - conf(X \Rightarrow Y) \quad (5)$$

The range of the effect metric is from -1 to 1 . Unlike the support, confidence, importance, and dominance that are all rule-wise metrics, effect is an item-wise metric.

C. SYSTEM USE CASES AND INTERFACE

The icuARM features a user interface that allows real-time association rule mining in the ICU. The interface enables the user to input real-time patient clinical scenarios, extract confident association rules from the database, and display the rules.

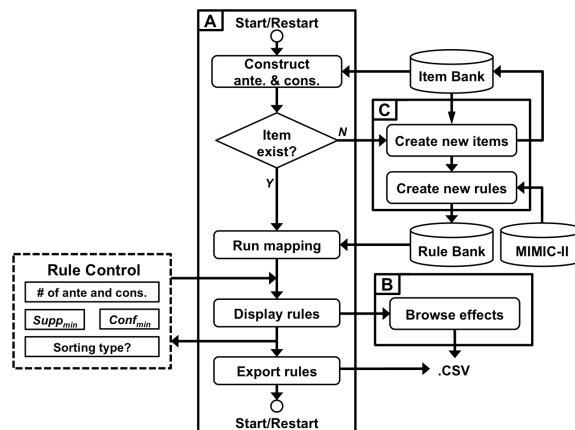


FIGURE 1. System Use Flow. The system consists of three main windows, including a Rule Mining window (A), an Effect Browsing window (B), and a New Item window (C). The Rule Mining window receives user inputs as Rule Control (dashed box). The system synchronizes with three sources, including the MIMIC-II database, Item Bank that stores all created items, and Rule Bank that stores all pre-existing rules.

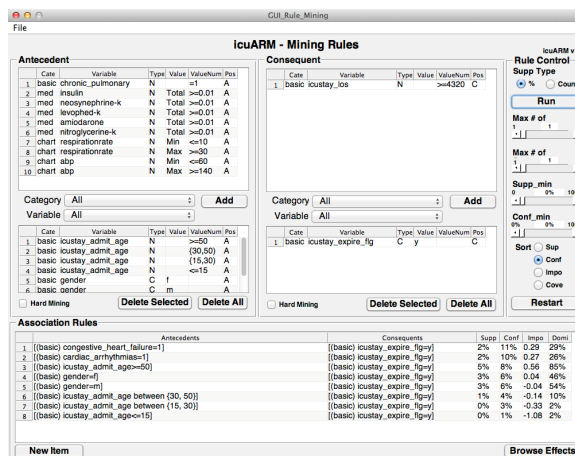


FIGURE 2. The Rule Mining Window. Users can construct antecedents and consequents of interest by selecting items in the Antecedent panel and the Consequent panel, respectively. The association rule results are displayed in the Association Rules panel. Users can manipulate rules by giving control inputs in the Rule Control panel. The control inputs include the type of $Supp_{min}$, maximum length of items in antecedents and consequents, $Supp_{min}$, $Conf_{min}$, and the sorting type. Rules can be exported from the "File" menu. Users can access the New Item window and Effect Browsing window via the bottom two buttons.

The icuARM usage flow is shown in Fig. 1. As depicted, the icuARM system consists of three main windows: (1) a Rule Mining window, (2) an Effect Browsing window, and (3) a New Item window. The interface was implemented in MATLAB (MathWorks, Natick, MA). The user can freely switch among these three windows for different purposes. The detailed features of each window are described below.

The first main window is the Rule Mining window (Fig. 2) that enables clinicians to extract association rules based on customized antecedents and consequents. The window accesses two data sources. The first data source of icuARM is the Rule Bank which stores rules that were previously mined. The second data source is the Item Bank that stores all

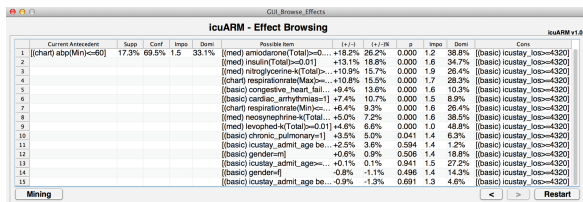


FIGURE 3. The Effect Browsing Window. Users can browse the effects of possible items on the right. The antecedent with selected items is shown on the left with the current rule measures. Users can keep selecting possible items from the right until no more potential items are available. Users can return to the Rule Mining window by clicking the Mining button.

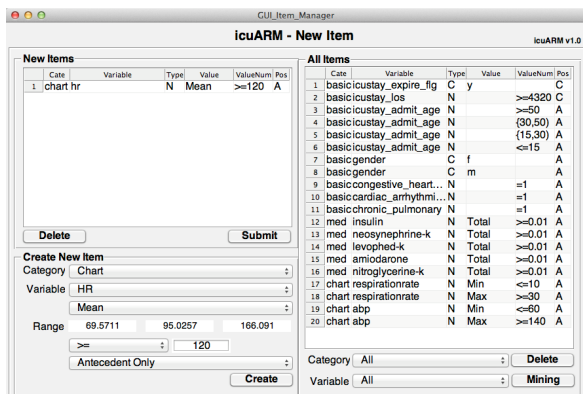


FIGURE 4. The New Item Window. In the Create New Item panel, users can create new items by selecting a variable under a category and assigning a value or a range of values. The pending new items are listed in the New Items panel. By clicking the Submit button, the system stores the new items and generates all corresponding rules. Users can browse all existing items in the All Items panel. Users can return to the Rule Mining window by clicking the Mining button.

pre-existing items. The mining process starts by constructing all items of interest in antecedents and consequents. Based on these items, the system retrieves all possible raw rules from the Rule Bank based on $Supp_{min} = 0\%$ and $Conf_{min} = 0\%$. The user can apply several inputs to prune out infrequent and/or unconfident rules by increasing $Supp_{min}$ and/or $Conf_{min}$, respectively. Because antecedents and consequents of the displayed rules may contain multiple items, the user can specify the length (i.e., the number of items) of processed antecedents and/or consequents. In addition, the rules can be sorted by one of the four rule-wise metrics (i.e., support, confidence, importance, and dominance). Furthermore, the user can export all raw rules or processed rules in comma-separated values (CSV) format, which can be exported into Microsoft Excel or other statistical analysis tools (e.g., SPSS) for future analysis.

The second main window is the Effect Browsing window (Fig. 3) that enables clinicians to browse item-wise effects after the rules are extracted from the Rule Bank. When given a target consequent, this interface starts by displaying all rules that contain only one item in their antecedents with their corresponding four rule-wise metrics. Once a first-item has been selected in the antecedent, the browser lists all possible second-items with their corresponding effects (i.e., the confidence changes). The browsing con-

tinues until no more potential items can be selected in the antecedent.

The third main window is the New Item window (Fig. 4) that enables clinicians to create new items. As mentioned in the Rule Mining window, the user can select items to construct customized antecedents and consequents. However, the user may not always find items of interest that have been created in the system. Therefore, the New Item window allows the user to construct new items by selecting a category, choosing a variable (i.e., the first component of the item) under the category, and assigning a value or a range of values (i.e., the second component of the item) to the variable. If the chosen variable is event-based, the user can also assign a value of mean, minimum, maximum, and standard deviation (as we mentioned in Section II.A). The user can also select the event duration if it is available. Once the user submits new items, the system stores them in the Item Bank, accesses the MIMIC-II database, and generates new rules. The efficiency of rule generation depends on the computation power of the running PC and can be expedited by parallel computing. The current system can generate from 52 rules per minute on a single-core processor to 434 rules per minute on a 12-core processor. Upon creation of new rules, the system stores them in the Rule Bank. Afterwards, users can retrieve the rules in the Rule Bank via the Rule Mining window without waiting for duplicated rule generation. As the number of stored rules keeps growing, a real-time rule mining system can gradually be achieved.

III. RESULTS AND DISCUSSION

A. PRE-EXISTING COMORBIDITY VS. PROLONGED ICU STAY

The length of stay (LOS) is a significant ICU outcome that is associated with severe organ failure and high resource consumption [23, 24]. Evidence has shown that patients with prolonged (i.e., longer than 3 days) ICU stays have a considerably increased ICU, hospital, and long-term mortality [25]. Patient comorbidity is a significant variable affecting the ICU LOS [26]. However, survival is typically estimated on a long-term basis (e.g., 1-yr or 2-yrs survival) that is not applicable to the short-term ICU prediction. Therefore, in this case study, we employed the icuARM to generate association rules between pre-existing comorbidities and prolonged ICU stays.

Table 2 lists the four metrics of rules for 12 possible pre-existing comorbidities. The rule with hypothyroidism (HYP) as the comorbidity was ignored because its rule had importance less than 1. Congestive heart failure (CHF) had the highest support (8.6%), which means that this rule was applicable to the highest portion of the ICU stays. Additionally, according to the dominance metric, CHF dominated the prolonged ICU stays by 22.4%, which was also the highest.

In this case study, the possibility of prolonged ICU stay can be predicted by the confidence of a rule given the comorbidities in the antecedent. According to the importance rules (i.e., importance ≥ 1) shown in Table 2, coagulopathy (COA) is associated with the highest possibility of prolonged ICU stay

TABLE 2. Pre-existing comorbidities VS. prolonged ICU stay.

Abbr.	Comorbidity	Supp (%)	Conf (%)	Impo	Domi (%)
COA	Coagulopathy	2.8	54.1	1.4	7.4
CHF	Congestive Heart Failure	8.6	49.9	1.4	22.4
CAA	Cardiac arrhythmias	7.3	46.0	1.3	19.1
REF	Renal failure	2.4	44.6	1.2	6.3
LID	Liver disease	1.9	44.5	1.2	5.0
OBE	Obesity	0.6	43.0	1.1	2.4
CHP	Chronic pulmonary	5.6	41.5	1.1	14.8
DEA	Deficiency anemia	3.6	40.6	1.1	9.4
DIA	Diabetes	6.9	38.9	1.1	18.6
CAN	Cancer	4.5	37.1	1.0	12.2
ABU	Alcohol/drug abuse	2.0	36.9	1.0	5.3
HYP	Hypothyroidism	2.3	35.0	0.9*	6.0

Supp, support; Conf, confidence; Impo, importance; Domi, dominance
Confidence is the possibility of prolonged ICU stay.

*Rule is ignored because the importance is less than 1.

(54.1%), whereas alcohol/drug abuse (ABU) is associated with the lowest possibility (36.9%).¹ The association rules of the eight age-gender populations (i.e., age <2, 15-30, 30-50, >50 years, and gender male and female) were also generated. Fig. 5(a) shows that females aged over 50 had the highest possibility of prolonged ICU stay (38.8%), and males over 50 years had the second highest (38.2%). Interestingly, the pediatric population had the third and fourth highest possibility of prolonged ICU stay (female: 37.6%, male: 38.1%). This may be a reflection of the fact that most of the children within the MIMIC-II database are neonates in the neonatal ICU, where premature infants tend to have prolonged stays (up to months).

The icuARM’s Effect Browsing window was used to investigate the effect of different combinations of pre-existing comorbidities in different populations on the possibility of prolonged ICU stay. We focused on females aged over 50 because of their highest possibility of prolonged ICU stay. Fig. 5(b) shows all rules of the 11 first-item comorbidities in this population. Coagulopathy (COA) was still associated with the highest possibility (58.6%) of prolonged ICU stay. In addition, females over 50 years who had alcohol and/or drug abuse showed an increased possibility (44.1%) even though this comorbidity did not have a high risk in the general population.

We continued to investigate the effects (i.e., changes of possibility of prolonged ICU stay) of possible second-item comorbidities in females aged over 50 years who also had coagulopathy. As shown in Fig. 5(c), there were 10 possible second-item comorbidities that were important (importance ≥ 1). Among them, six comorbidities increased the possibility, with deficiency anemia (DEA) resulting in the highest rate of prolonged ICU stay (64.4%). Clinicians can continue the effect browsing process by adding other comorbidity combinations based on a patient’s status at admission.

¹ Some comorbidities tend to be associated with shorter ICU stays, especially in younger population. For example, young patients with ABU usually admit for short (i.e., < 24 hrs) ICU stays because of acute alcohol intoxication (e.g., seizures/delirium tremens) or drug overdose/intoxication (e.g., respiratory monitoring after an opiate or benzodiazepine overdose).

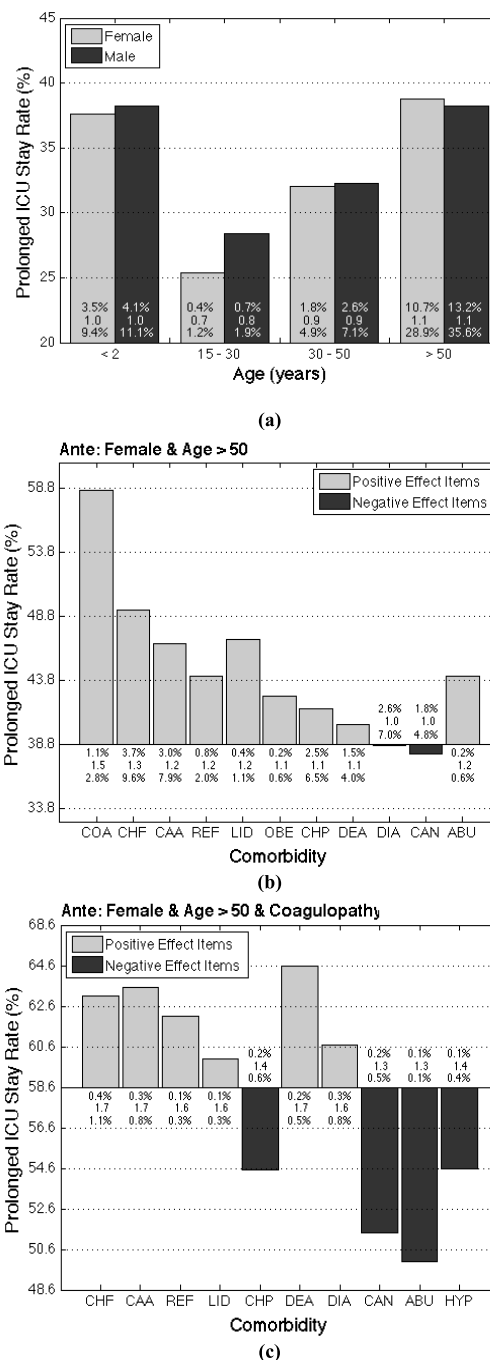


FIGURE 5. (a) Possibility of prolonged ICU stays in different age-gender populations. The three values of each bar are the three measures of the association rules, including support (top), importance (middle), and dominance (bottom). (b) The effects (i.e., changes in possibility of prolonged ICU stay) of the first-item comorbidities on prolonged ICU stay possibility in females aged over 50 years. (c) The effects of the second-item comorbidities on prolonged ICU stay possibility in females aged over 50 years who also have coagulopathy.

After evaluating rules based on their support and confidence values, it is important to interpret the importance value. Rules have lower importance values for more general cases with fewer items (in either the antecedent or consequent), and have higher importance values for more specific cases with more items. When comparing the importance values among

Table 2 and Figure 5(a) to 5(c), we can observe that the importance values increase as more items are added to the antecedents. We want to emphasize 1-item rules in Table 2. So even if, theoretically, the value can go to infinity, in reality, it is more around 1.1-1.4.

In this case study, we have shown the basic usability of icuARM to assess associations between pre-existing comorbidities and prolonged ICU stays, especially in females aged over 50 years. Clinicians can construct different combinations of age, gender, and pre-existing comorbidities to determine a baseline prolonged ICU stay possibility of a patient at the time of ICU admission, even prior to diagnosis. By estimating the possibility of prolonged ICU stay, an ICU team can efficiently plan ahead for the intensive care resource allocation such as staffing, laboratory, and radiology. This prediction also provides a risk reference to assess how certain interventions will affect LOS. For example, a clinician may admit two female patients of similar age. One has a coagulopathy (e.g., disseminated intravascular coagulation) and the other has an acute coronary syndrome. By using icuARM, the clinician and ICU team could accurately plan for needed resources for the former patient, predict outcomes, and improve management for this type of high risk ICU patient.

B. MEDICATION USAGE VS. PROLONGED ICU STAY

Mining associations between medication usage and clinical outcome is another promising application of icuARM. ARM has been adopted in several pharmacovigilance studies, such as investigating multi-item adverse drug reactions [27-29]. However, to our knowledge, no CDSSs have adopted ARM for finding associations between medication usage and ICU outcomes. Therefore, in our second case study, by using icuARM, we investigated the associations between prolonged ICU stays and medication usage in addition to patient demographics and pre-existing comorbidities.

We first mined the association rules of two commonly used anti-hypertensive drugs in ICUs: diltiazem (DIL) and labetalol (LAB). We selected males and females over 50 years because they had the highest prolonged ICU possibility according to our previous case study. The associations on the drugs with a pre-existing comorbidity of congestive heart failure (CHF) were also investigated. As shown in Table 3, in patients over 50 years without CHF, DIL is associated with higher possibility compared to LAB in both females and males. However, these two drugs had different effects on patients with CHF. For females over 50 years with CHF, the use of DIL increased the possibility of prolonged ICU stays to 83.4% compared to females over 50 without CHF (73.6%), whereas LAB had nearly no change (62.3% vs. 61.7%). In contrast, for the same clinical situation, the use of LAB actually increased the possibility in males over 50 years with CHF to 87.1% compared to those without CHF (62.8%), whereas DIL had almost no effect (74.7% vs. 76.0%). Therefore, for patients over 50 years with a comorbidity of CHF, we may choose LAB for females and DIL for males.

TABLE 3. Medication usage Vs. prolonged ICU stay (>50 years old).

Comorbidity	Gender	Probability of Prolonged ICU Stay (%)				
		LAB	DIL	EPI	VAS	EPI+VAS
No CHF	Female	62.3	73.6	67.6	64.7	67.6
	Male	62.8	74.7	61.8	71.4	74.0
Has CHF	Female	61.7	83.4	84.5	82.0	84.2
	Male	87.1	76.0	68.1	85.2	87.8

CHF, congestive heart failure; LAB, labetalol; DIL, diltiazem; EPI, epinephrine; VAS, vasopressin

In addition to the hypertensive conditions, ICU clinicians often have a choice between pharmacologic agents in an acute episode of cardiopulmonary arrest. Epinephrine (EPI) and vasopressin (VAS) are two common drugs used in the management of ventricular fibrillation and pulseless electrical activity. We applied icuARM to explore the associations between these two drugs and prolonged ICU stays. In addition, Gueugniaud *et al.* suggested that the combination of EPI and VAS did not improve outcome (i.e., survival to hospital discharge, good neurologic recovery, and 1-year survival) during advanced cardiac life support for out-of-hospital cardiac arrest [30]. However, evidence was still insufficient to make prognosis on short-term ICU stays. Therefore, by utilizing icuARM, the association between ICU LOS and a combination of EPI and VAS was also evaluated compared to EPI or VAS alone.

According to the result shown in Table 3, females over 50 years without CHF had slightly lower possibility of prolonged ICU stay with VAS compared to EPI (64.7% vs. 67.6%); in contrast, males over 50 without CHF had lower chance of prolonged ICU stay with EPI compared to VAS (61.8% vs. 71.4%). These associations all increased on patients over 50 who also had CHF, but the EPI increased the possibility most on females (84.5%) compared to those without CHF (67.6%). Furthermore, for the combination of EPI and VAS, the change of the possibility was not considerably different compared to EPI or VAS alone. This partially supported the finding of [30] although we focused on the short term ICU outcome.

This case study demonstrated that icuARM could help guide the clinician to select correct medication for similar clinical situations but different patient populations in the pre-planning phase. The entire mining process requires no more than one minute, promising a nearly real-time and easy-to-access bedside consulting tool.

IV. CONCLUSION

Evidence-based real-time decision-making for critically ill patients in the ICU has become more challenging because the volume and complexity of the data have been increasing over the years. Thus, to assist clinicians in making optimal decisions, there is a critical need to apply modern information technology and advanced data analytics to extract information from heterogeneous clinical data.

In this study, we researched and developed a real-time clinical decision support system icuARM to assist clinicians

in generating quantitative and real-time decision support rules for the ICU based on a large ICU patient database MIMIC-II. We adopted the “support” and the “confidence” metrics suitable for ICU clinical application from conventional association rule mining. In addition, we defined and developed two new rule-wise metrics “importance” and “dominance” and one item-wise metric “effect.” We developed an interactive and easy-to-use graphical user interface that enables clinicians to perform flexible data mining in real-time for personalized decision-making.

We tested icuARM on two cases investigating the associations between prolonged ICU stays and patient demographics, pre-existing comorbidities, and medication usage. Our results not only reinforced the current decision-making evidence, but also revealed new knowledge by predicting characteristics of a prolonged ICU stay.

We will further improve this CDSS in four directions. First, besides basic patient demographics, pre-existing comorbidity data, and medication usage, we will emulate more categories such as nurse-verified chart events, laboratory tests, and fluid balance records etc. to better assist ICU clinicians in making critical decisions. Second, as mentioned in Section II.A, we aim to include continuous physiological data with corresponding time stamps from the MIMIC-II database to perform temporal association rule mining [31]. Third, the current mining process with the Apriori algorithm requires clinicians to manually specify variables of interest and cut-points (for numerical variables) in the items of antecedents and consequents. We will develop automatic feature selection, such as the supervised mRMR method [32] or the unsupervised MCFS method [33], and discretization for more objective item construction. Finally, in addition to the five evaluation metrics, we will provide a more comprehensive clinical evaluation by including other data centric rule metrics [34].

ACKNOWLEDGMENT

The authors are grateful to Dr. John Phan, Chanchala Kaddi, and Po-Yen Wu for their valuable comments and suggestions.

REFERENCES

- [1] (April 2013). *The 2012 Statistics Brochure, Society of Critical Care Medicine* [Online]. Available: http://www.sccm.org/Public_Health_and_Policy/Pages/Statistics-Brochure.aspx
- [2] D. J. Cullen, B. J. Sweitzer, D. W. Bates, E. Burdick, A. Edmondson, and L. L. Leape, “Preventable adverse drug events in hospitalized patients: A comparative study of intensive care and general care units,” *Critical Care Med.*, vol. 25, no. 8, pp. 1289–1297, 1997.
- [3] L. B. Andrews, C. Stocking, T. Krizek, L. Gottlieb, C. Krizek, T. Vargish, et al., “An alternative strategy for studying adverse events in medical care,” *Lancet*, vol. 349, pp. 309–313, Feb. 1997.
- [4] J. Wyatt and D. Spiegelhalter, “Field trials of medical decision-aids: Potential problems and solutions,” in *Proc. Annu. Symp. Comput. Appl. Med. Care*, 1991, pp. 3–7.
- [5] D. L. Sackett, “Evidence-based medicine,” *Seminars Perinatol.*, vol. 21, no. 1, pp. 3–5, 1997.
- [6] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, et al., “Clinical decision support systems for the practice of evidence-based medicine,” *J. Amer. Med. Inf. Assoc.*, vol. 8, no. 6, pp. 527–534, 2001.
- [7] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman, “APACHE II: A severity of disease classification system,” *Critical Care Med.*, vol. 13, no. 10, pp. 818–829, 1985.
- [8] M. Verduijn, N. Peek, F. Voorbraak, E. De Jonge, and B. de Mol, “Dichotomization of ICU length of stay based on model calibration,” in *Artificial Intelligence in Medicine*. New York, NY, USA: Springer-Verlag, 2005, pp. 67–76.
- [9] F. L. Ferreira, D. P. Bota, A. Bross, C. Mélot, and J.-L. Vincent, “Serial evaluation of the SOFA score to predict outcome in critically ill patients,” *J. Amer. Med. Assoc.*, vol. 286, no. 14, pp. 1754–1758, 2001.
- [10] G. Teasdale and B. Jennett, “Assessment of coma and impaired consciousness: A practical scale,” *Lancet*, vol. 304, pp. 81–84, Jul. 1974.
- [11] K. A. Fox, O. H. Dabbous, R. J. Goldberg, K. S. Pieper, K. A. Eagle, F. Van de Werf, et al., “Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: Prospective multinational observational study (GRACE),” *BMJ*, vol. 333, no. 7578, pp. 1091–1096, 2006.
- [12] E. Turban, J. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems*, 7th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.
- [13] J. Ramon, D. Fierens, F. Güiza, G. Meyfroidt, H. Blockeel, M. Bruynooghe, et al., “Mining data from intensive care patients,” *Adv. Eng. Informat.*, vol. 21, no. 3, pp. 243–256, 2007.
- [14] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, et al., “Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access intensive care unit database,” *Critical Care Med.*, vol. 39, no. 5, pp. 952–960, 2011.
- [15] R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases,” in *Proc. ACM SIGMOD Rec.*, 1993, pp. 207–216.
- [16] S. Konias, G. Giaglis, G. Gogou, P. Bamidis, and N. Maglaveras, “Uncertainty rule generation on a home care database of heart failure patients,” in *Proc. Comput. Cardiol.*, 2003, pp. 765–768.
- [17] C. Ordonez, E. Omiecinski, L. de Braal, C. A. Santana, N. Ezquerria, J. A. Taboada, et al., “Mining constrained association rules to predict heart disease,” in *Proc. IEEE ICDM*, Nov./Dec. 2001, pp. 433–440.
- [18] Y. Shan, D. Jeacocke, D. W. Murray, and A. Sutinen, “Mining medical specialist billing patterns for health service management,” in *Proc. 7th Australasian Data Mining Conf.*, vol. 87, 2008, pp. 105–110.
- [19] R. Bellazzi, C. Larizza, P. Magni, and R. Bellazzi, “Temporal data mining for the quality assessment of hemodialysis services,” *Artif. Intell. Med.*, vol. 34, no. 1, pp. 25–39, 2005.
- [20] R. Chaves, J. Górriz, J. Ramírez, I. Illán, D. Salas-Gonzalez, and M. Gómez-Río, “Efficient mining of association rules for the early diagnosis of Alzheimer’s disease,” *Phys. Med. Biol.*, vol. 56, no. 18, pp. 6047–6063, 2011.
- [21] C.-L. Chang, “A study of applying data mining to early intervention for developmentally-delayed children,” *Expert Syst. Appl.*, vol. 33, no. 2, pp. 407–412, 2007.
- [22] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in *Proc. 20th Int. Conf. VLDB*, 1994, pp. 487–499.
- [23] H. Kern, U. Redlich, H. Hotz, C. Von Heymann, J. Grosse, W. Konertz, et al., “Risk factors for prolonged ventilation after cardiac surgery using APACHE II, SAPS II, and TISS: Comparison of three different models,” *Intensive Care Med.*, vol. 27, no. 2, pp. 407–415, 2001.
- [24] W. A. Knaus, D. P. Wagner, J. E. Zimmerman, and E. A. Draper, “Variations in mortality and length of stay in intensive care units,” *Ann. Internal Med.*, vol. 118, pp. 753–761, May 1993.
- [25] O. V. Hein, J. Birnbaum, K. Wernecke, M. England, W. Konertz, and C. Spies, “Prolonged intensive care unit stay in cardiac surgery: Risk factors and long-term-survival,” *Ann. Thoracic Surgery*, vol. 81, no. 3, pp. 880–885, 2006.
- [26] D. C. Angus, W. T. Linde-Zwirble, C. A. Sirio, A. J. Rotondi, L. Chelluri, R. C. Newbold, et al., “The effect of managed care on ICU length of stay: Implications for medicare,” *J. Amer. Med. Assoc.*, vol. 276, no. 13, pp. 1075–1082, 1996.

- [27] J. Chen, H. He, J. Li, H. Jin, D. McAullay, G. Williams, et al., "Representing association classification rules mined from health data," in *Proc. 9th Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, 2005, pp. 1225–1231.
- [28] R. Harpaz, H. Chase, and C. Friedman, "Mining multi-item drug adverse effect associations in spontaneous reporting systems," *BMC Bioinform.*, vol. 11, pp. S7–S14, Oct. 2010.
- [29] M. Rouane-Hacene, Y. Toussaint, and P. Valtchev, "Mining safety signals in spontaneous reports database using concept analysis," in *Artificial Intelligence in Medicine*. New York, NY, USA: Springer-Verlag, 2009, pp. 285–294.
- [30] P.-Y. Gueugniaud, J.-S. David, E. Chanzy, H. Hubert, P.-Y. Dubien, P. Mauriacourt, et al., "Vasopressin and epinephrine vs. epinephrine alone in cardiopulmonary resuscitation," *New England J. Med.*, vol. 359, pp. 21–30, Jul. 2008.
- [31] E. Winarko and J. F. Roddick, "ARMADA—An algorithm for discovering richer relative temporal association rules from interval-based data," *Data Knowl. Eng.*, vol. 63, no. 1, pp. 76–90, 2007.
- [32] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [33] D. Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 333–342.
- [34] P.-N. Tan, V. Kumar, and J. Srivastava, "Selecting the right objective measure for association analysis," *Inf. Syst.*, vol. 29, no. 4, pp. 293–313, 2004.



CHIH-WEN CHENG (M'09) received a B.S. (Bachelor of Science) degree in Electrical Engineering from Chung Cheng University, Chia-Yi, Taiwan, in 2005, and a M.S. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, in 2009.

He is currently pursuing his Ph.D. as a graduate research assistant in the School of Electrical and Computer Engineering. His research focuses on applying data mining, cloud computing, and mobile technologies to improve clinical decision-making to facilitate quality-of-healthcare. He is actively involved in projects collaborating with clinical and research institutes from the metro-Atlanta area, including Children's Healthcare of Atlanta, Georgia State University, and Shepherd Center. His tools and publications are applicable to areas of sickle cell disease, acute heart failure, pediatric psychological disorder, and rehabilitation in traumatic brain injury.



NIKHIL CHANANI received the B.S. degree in biological sciences and the B.A. degree in economics from Stanford University, Palo Alto, CA, USA, in 1997, and the M.D. degree from Harvard Medical School, Boston, MA, USA, in 2002. He completed the postgraduate training in pediatrics and pediatric cardiology with the University of California, San Francisco, CA, USA, from 2002 to 2008 and additional training in Pediatric Cardiac Intensive Care at Lucile Packard Children's Hospital, Stanford University from 2008 to 2010.

He is currently an Assistant Professor of pediatrics with the Emory University School of Medicine and an attending physician in the Cardiac Intensive Care Unit, Children's Healthcare of Atlanta, Atlanta, GA, USA. He is a Staff Cardiologist with the Sibley Heart Center. His research interests include quality and device improvement for patients with congenital heart disease.

Dr. Chanani is a member of the American Heart Association, the American College of Cardiology, and the Pediatric Cardiac Intensive Care Society.



JANANI VENUGOPALAN (M'12) received the B.Tech. degree in biomedical engineering from Sathyabama University, Chennai, India, in 2008, the B.E. degree in computer science from Sathyabama University in 2009, the M.Tech. degree in clinical engineering from the Indian Institute of Technology, Madras, Chennai, in 2011.

She was a Systems Design Engineer with Perfint Healthcare, and interned in Christian Medical College and All India Institute of Medical Sciences.

She is currently pursuing the Ph.D. degree in biomedical engineering with the Georgia Institute of Technology, Atlanta, GA, USA. Her research focuses on mobile health and health informatics. She is actively involved in projects pertaining to clinical decision making using EHR and PHR, which are currently in collaboration with Children's Healthcare of Atlanta and Shepherd Center. Her tools and publications find applications in sickle cell disease, traumatic brain injury, and acute heart failure.



KEVIN MAHER received the B.S. degree in biological sciences from Virginia Tech, Blacksburg, VA, USA, in 1984, and the M.D. degree from the University of Maryland, Baltimore, MD, USA, in 1991. He was with the University of Maryland for training in pediatrics and then spent a year as Chief Resident of pediatrics. He was with the University of Michigan for a fellowship in pediatric cardiology. He was involved in cardiac intensive care and interventional catheterization at Thomas Jefferson

University, DuPont Children's Hospital. He was involved in pediatric cardiac intensive care at the Children's Healthcare of Atlanta, Emory University School of Medicine, in 2004, where he is an Associate Professor of pediatrics. His research interests are varied and include neonatal resuscitation, biomarkers in pediatric heart disease, and the application of advanced technologies to pediatric disease. He is a founding member and co-director of the Atlanta Pediatric Device Consortium, The Center for Pediatric Innovation, and the Center for Pediatric Nanomedicine. Each of these programs represents collaborative efforts between Children's Healthcare of Atlanta, Georgia Tech, and Emory University. He has participated in numerous international medical programs to provide pediatric cardiac care in underserved populations. He is a National Board Member for the American Heart Association and is a member of the American College of Cardiology and the Pediatric Cardiac Intensive Care Society.



MAY DONGMEI WANG (SM'11) received a B.S. degree from Tsinghua University, Beijing, China, M.S. Electrical and Computer Engineering, M.S. Applied Mathematics, M.S. Computer Science, and a Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, USA. She has industrial R&D experience in the former AT&T Bell Labs, Lucent Technologies Bell Labs, Intel Architecture Lab, and Hughes Research Labs. Dr. Wang is currently

a tenured Associate Professor in The Joint Wallace H. Coulter Department of Biomedical Engineering, School of Electrical and Computer Engineering, Department of Hematology and Oncology, The Winship Cancer Institute, The Parker H. Petit Institute for Bioengineering and Biosciences, and The Institute for People and Technology at Georgia Institute of Technology and Emory University, Atlanta. Dr. Wang is a Georgia Cancer Coalition Distinguished Cancer Scholar, Director of the Biocomputing and Bioinformatics Core in the Emory-Georgia Tech Cancer Nanotechnology Center, and Co-Director of the Georgia Tech Center of Bio-Imaging Mass Spectrometry. Her research focuses on Biomedical and Health Informatics for personalized health, including high throughput -omic data analysis for clinical biomarker identification, tissue imaging informatics, systems modeling for predictive medicine, and health informatics. Dr. Wang has received Georgia Tech Outstanding Faculty Mentor Award for Undergraduate Research in 2005. She has been serving as the co-chair for Biomedical and Health Informatics Technical Committee of IEEE EMBS since 2011.