International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:7, 2012

Clinical Decision Support for Disease Classification based on the Tests Association

Sung Ho Ha, Seong Hyeon Joo, Eun Kyung Kwon

***Abstract*—**Until recently, researchers have developed various tools and methodologies for effective clinical decision-making. Among those decisions, chest pain diseases have been one of important diagnostic issues especially in an emergency department. To improve the ability of physicians in diagnosis, many researchers have developed diagnosis intelligence by using machine learning and data mining. However, most of the conventional methodologies have been generally based on a single classifier for disease classification and prediction, which shows moderate performance. This study utilizes an ensemble strategy to combine multiple different classifiers to help physicians diagnose chest pain diseases more accurately than ever. Specifically the ensemble strategy is applied by using the integration of decision trees, neural networks, and support vector machines. The ensemble models are applied to real-world emergency data. This study shows that the performance of the ensemble models is superior to each of single classifiers.

***Keywords*—**Diagnosis intelligence, ensemble approach, data mining, emergency department

I. Introduction

THE hospital's emergency department is a complex unit in which the fight between life and death is only a breath away. The emergency department has been frustrated by the problems of overcrowding, long waiting time, patient care delays, and high costs over decades. Accordingly, to solve these problems has become the hottest issue in this area. Several internal or external factors have contributed to the long processing time and patient care delay: patient characteristics, emergency department staffing patterns, access to health care providers, patient arrival time, management practices, and testing and treatment strategies chosen [12]. Understanding these factors well is an important step to improve the efficiency of patient care in an emergency department.

Most hospitals today have employed certain kinds of hospital information systems to manage their healthcare or patient data. These systems typically generate vast amounts of data in the forms of number, text, chart, and image. This raises an important question: "How can healthcare practitioners turn that data into useful information that would enable to make intelligent clinical decisions?" Considering the fast growth of data content, size, and diversity, researchers have focused on techniques to find useful information from collections of data during previous decades. Although its application to medical data analysis has been relatively limited until recently, the term 'data mining' has been increasingly used in the medical literature [11]. The goal of predictive data mining in clinical medicine is to derive models that use specific patient information to support clinical decision-making.

Professor Sung Ho Ha is with School of Business Administration, Kyungpook National University, 1370 Sangyeok-dong, Buk-gu, Daegu, Korea. (e-mail: hsh@mail.knu.ac.kr)

PhD Seong Hyeon Joo is with School of Business Administration, Kyungpook National University, Korea. (e-mail: sideas@mail.knu.ac.kr)

PhD Eun Kyung Kwon is with School of Business Administration, Kyungpook National University, Korea. (e-mail: 01045332694@nate.com)

Data mining models can be applied to building of decision-making procedures such as prognosis, diagnosis, and treatment planning, which once evaluated and verified, could then be embedded within clinical information systems [31].

Accordingly, the purpose of this study is as follows: first, using data mining techniques, this study focuses on generating the association rules that help physicians to decide which lab tests they should select, which can reduce lab-testing time and cost in the emergency department. Second, this study aims to build an ensemble of classifiers that supports to make a complex diagnosis, which can help physicians to formulate clinical decisions more quickly and more accurately. The organization of the paper is as follows: Section 2 explains medical data mining appeared in the literature and its application to the emergency department. Section 3 illustrates a hybrid decision model as a research methodology used in this study and section 4 applies the methodology to the real emergency data. Section 5 evaluates the methodology and compares it with other techniques. Section 6 provides conclusions and future directions.

II. Literature Review

A. Medical data mining

Medical data mining has been applied to accurate classification and rapid prediction for prognosis and diagnosis of patients in a specialized medical area [5][24]. It has been also used for training unspecialized doctors to solve a specific diagnostic problem [20]. Among several algorithms for classification and prediction tasks, a decision tree is one of the frequently used techniques in medical data mining area [15]. While it is easy to find many cases to prove the decision tree to be useful in the business domain, the decision tree enables to predict prognoses and diagnoses in the domain of medicine, by using tree-structured models or in the form of 'IF condition-based-on attribute-values THEN outcome-value' to identify useful features of importance. Khan et al. [19] used decision trees to extract clinical reasoning in the form of medical expert's actions that are inherent in a large number of electronic medical records. The extracted data could be used to teach students of oral medicine a number of orderly processes for dealing with patients with different problems, depending on the time. Yun [38] utilized a C4.5 algorithm to build a decision tree in order to discover the critical causes of type II diabetes. She has learned knowledge about the illness regularity from diabetes data, and has generated a set of rules for diabetes diagnosis and prediction. The Apriori algorithm was useful to figure out large item sets and thus to generate association rules from medical data. Abdullah et al. [1] adopted an association algorithm to find the relationship between diagnosis and prescription. They stated that purchases and medical bills had much in common. Tan et al. [36] used the Apriori algorithm to mine the rules for the compatibility of drugs from prescriptions to cure arrhythmia in the traditional Chinese medicine database.

The experimental results showed that the drug compatibility obtained by the Apriori algorithm generally was consistent with the traditional Chinese medicine for that disease. Emergency data is the data collected by emergency department environment. They are more critical to human life than routine medical data. Most diseases included in the emergency data are fatal diseases. Ceglowski et al. [4] discovered 'treatment pathways' through mining medical treatment procedures in the emergency department. They found that the workload in the emergency department varied depending on the number of presented patients, and was not affected by the type of procedure carried out. Rossille et al. [33] have presented a complementary perspective on the activities for specific patient groups by the emergency department: patients over 75 year old or less than 75 year old. She thought once validated, these views would be used as decision support tools for delivering better care to this population. Lin et al. [21] found a way to raise the accuracy of triage through mining abnormal diagnostic practices in the triage. A two-stage cluster analysis (Ward's method, K-means) and a decision tree analysis were done on abnormal diagnoses in an emergency department.

Artificial neural networks have been applied to the fields of clinical diagnosis mostly to render the complex and fuzzy cognitive process of diagnosis. Many previous studies have shown the suitability of neural networks in the design of clinical decision support systems and biomedical applications. Ellenius et al., [9] assessed chest-pain patients and declared the limits of critically-sized systematic errors by calculating the decrease in diagnostic performance of neural networks. Guven and Kara [14] concentrated on the diagnosis of subnormal eye through the analysis of Electrooculography signals with the help of neural networks. Support vector machines (SVM) are recently of increasing interest to biomedical researchers. It is because of not only the characteristics of well-founded theory, but also the superiority in practical applications. Conforti and Guido [6] developed kernel-based SVM classifiers to aid the early diagnosis of acute myocardial infarction. By running a 10-fold cross validation procedure, the performance of their classifier was 97.5%. Majumder et al. [22] made use of SVM for classification by integrating the recursive feature elimination.

B. Chest pain research

All tables and figures you insert in your document are only to help you gauge the size of your paper, for the convenience of the referees, and to make it easy for you to distribute preprints. Chest pain is one of the most common reasons why people visit emergency rooms. Chest pain is particularly important because it may announce the existence of a serious and occasionally life-threatening disease. In addition, it can be complicated by the frequent disassociation between signal strength (symptoms) and seriousness of underlying pathology.

Considering the ratio of the outbreak of chest pain resulting from each body organ, it is easy to find that cardiac diseases are the most common cause of chest pain (45%), followed by musculoskeletal (14%), psychiatric (8%), gastrointestinal (6%), and pulmonary (5%) [27]. In fact, cardiac diseases are the third most common diseases in Korea (22,347 people died in 2009) and they have maintained the status over the last ten years [35].

Angina pectoris (AP) and acute myocardial infarction (AMI) are two types of acute illness, which can easily lead to death.An accurate diagnosis and treatment of patients with chest pain is a difficult task of emergency physicians. A procedure to diagnose patients with acute chest pain should identify high-risk patients quickly for a fast cure [10]. Bassan et al. [2] evaluated the efficiency of a systematic diagnostic approach to chest pain patients in an emergency room in the relationship between diagnosis of acute coronary syndrome and hospitalization rates in the high-cost care units. Martinez-Selles et al. [23] described the characteristics of patients with chest pain and evaluated the usefulness of the CPU-65 index for risk stratification of chest pain. Their research showed that the most frequent diagnosis for patients discharged from the emergency department was atypical chest pain and respiratory infection.

Ridker et al. [32] examined a group of healthy postmenopausal women over three years to assess the risk of cardiovascular events associated with base-line levels of markers of inflammation. Conroy et al. [7] initiated a Systematic Coronary Risk Evaluation project to develop a risk scoring system for use in the clinical management of cardiovascular risk in European clinical practice. Velagaleti et al. [37] found the relation of lipid concentrations to heart failure, taking into account age, gender, and associated cardiovascular risk factors.

III. Research Methodology

A. Emergency department process

There have been many studies that focus on redesigning and enhancing efficiency of the emergency department process. Based on these studies, a common process of current emergency department can thus be depicted. A regular patient enters the emergency department, picks a number, and remains in the waiting area. When their number is called, the patient is assessed by a triage nurse who screens for apparent critical symptoms. If the patient is found to be in critical condition, they are transferred to the Intensive Care Unit for immediate care. Otherwise, the triage nurse assigns a triage code depending on the patient's condition (1 to 5, 1 being most critical).

Once a triage code is assigned, the patient waits for a physician's assessment. The waiting period depends on the availability of physicians and examination rooms. After the first assessment, lab tests may be required by the physician (blood test, lung scanner, and so on). If not, the patient is discharged to go home or transferred to another department. A patient with complete lab test results has to wait again for a second assessment by the physician requesting the tests. After the second assessment, the patient may be discharged to go home with a prescription or transferred for admission. Patients arriving by ambulance are transferred directly to the trauma room without triage [8].

B. Diagnosis intelligence using data mining

Based on the process in an emergency department, this study concentrates on the steps between 'Lab tests required' → 'Tests completed' → 'Waiting' → 'Physician assessment'.

International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:7, 2012

They are where the problems of congestion, long waiting time occur seriously over decades in the emergency department. If some unnecessary tests can be removed without any loss of diagnostic accuracy, the cost caused from congestion and long waiting time can be reduced.

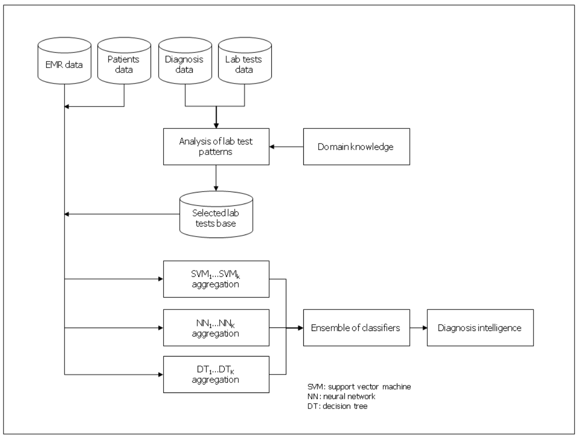

Fig. 1 A research methodology for diagnosis intelligence using an ensemble of classifiers

Fig. 1 illustrates the research methodology, which consists of three stages, such as data collection, feature selection, and an ensemble strategy. The first stage of the methodology is collecting data in the form of electronic medical record (EMR) from an emergency department of a hospital. The EMR records include the information of patients, lab tests, and diagnosis.

Feature selection is the second stage of importance in this study. It reduces the dimension of EMR data in such a way that selected features can represent the most significant aspects of data. This study reduces the number of lab tests necessary to diagnose chest diseases accurately. In doing so, this study extracts association relationships between lab tests and diagnosis, which are revealed by the Apriori algorithm in the form of association rules. The association rule here is an implication of the form $X \rightarrow Y$, which means 'If a patient takes a lab test X, then he will take a lab test Y'. The rule $X \rightarrow Y$ has to satisfy pre-specified minimum support and minimum confidence levels.

Domain knowledge about diagnostic tests for the chest pain diseases helps select the lab tests associated with those diseases. Thus, the second stage includes such lab tests, which domain knowledge mentions they are important, from a set of lab tests chosen from the association rules that have recorded higher scores on support and confidence levels (e.g., above 0.9). For example, if X is one of the critical tests mentioned in the domain knowledge, then, a test Y will be selected from the rules in the form of $X \rightarrow Y$ or $Y \rightarrow X$, whose support and confidence values are higher than 0.9.

In the third stage, by using the lab tests selected in the second stage and other information of patients (e.g., previous diagnosis and medical records), an ensemble strategy of classification algorithms is adopted to classify diagnosis of chest pain. Several different types of classification algorithms, including decision trees (DT), neural networks (NN), and support vector machines (SVM) are chosen to use.

Specifically, DT algorithms involve C5.0 (Entropy index) and Classification and regression tree (Gini and Twoing indexes). NN algorithms include Multilayer perceptron (MLP), Radial basis function network (RBFN), and Exhaustive prune. SVM algorithms involve RBF and Polynomial kernels.

After generating each single classifier, it needs to aggregate independent trained classifiers into DT ensembles, NN ensembles, and SVM ensembles through using an appropriate aggregation method, voting and confidence-weighted voting. Furthermore, the ensembles with higher performance can be collected to generate the final ensemble model. The one with the best performance will be the best final ensemble model.

1. Feature selection from lab tests

The association algorithm finds the associative relationship between lab tests. Whenever a patient gets medical lab tests, associations occur between lab tests regarding diagnosis. The Apriori algorithm is chosen to analyze the associations and select the most critical features.

The information of lab tests and diagnosis is extracted from the EMR, and then the chest pain diseases involved in the data are clustered into several kinds, i.e., acute myocardial infarction, angina pectoris, and so on. The clustered lab test data according to each disease is analyzed by the Apriori algorithm that generates association relationships between lab tests. The frequent association rules that contain critical lab tests revealed by domain experts are selected from the generated association rules, which form a knowledge base called 'Lab tests association base'.

Association rules mining using Apriori is a two-step process, where in the first step frequent item-sets (in this case, lab tests) are discovered and in the second step, association rules are derived from the frequent item-sets [17][26]. For the first step, the algorithm makes multiple scans over the data. In the first pass, the support of individual items is counted and frequent items are determined. In each subsequent pass, a set of item-sets found to be frequent in the previous pass is used for generating new frequent item-sets, called candidate item-sets, and their actual support is counted during the pass over the data. At the end of the pass, those satisfying minimum support constraint are collected, and they become the seed for the next pass. This process is repeated until no new frequent item-sets are found.

The second step is to generate the desired association rules from the frequent item-sets (in this case, lab tests). All subsets of every frequent item-set f are enumerated and for every such subset a, a rule of the form a $\rightarrow$ (f – a) is generated if the ratio of support(f) to support(a) is at least minimum confidence. Given a set of transactions $\{D = T_1, \ldots, T_n\}$ and a set of items $\{I = I_1, \ldots, I_m\}$, such that any transaction T in D is a set of items in I, an association rule is an implication $A \rightarrow B$ where the antecedent A and the consequent B are subsets of a transaction T in D, and A and B have no common items. Statistical interestingness can be measured according to various criteria, which are related to the observed frequency of the rules [39]. The support for a rule $A \rightarrow B$ is obtained by dividing the number of transactions, which satisfy the rule, $N_{A \rightarrow B}$, by the total number of transactions, N:

$$\text{Support}(A \rightarrow B) = \frac{N_{A->B}}{N} \quad (1)$$

The confidence of the rule $A \rightarrow B$ is obtained by dividing the number of transactions, which satisfy the rule by the number of transactions, which contain the body of the rule, *A*:

$$\text{Confidence}(A \rightarrow B) = \frac{N_{A->B}}{N_A} = \text{support}(A \rightarrow B) / \text{support}(A) \quad (2)$$

The lift takes the confidence of a rule and relates it to the support for the rule's consequent:

$$\begin{aligned}\text{Lift}(A \rightarrow B) &= \text{confidence}(A \rightarrow B) / \text{support}(B) \\ &= \text{support}(A \rightarrow B) / \text{support}(A)\text{support}(B) \quad (3)\end{aligned}$$

Notice that a lift value greater than 1 indicates there is a positive association, whereas a value less than 1 indicates there is a negative association.

2. Ensemble strategy for clinical classification

With the selected lab tests discovered by the 'Feature selection from lab tests' stage and other information (diagnosis, patient, and medical record data), an ensemble strategy with several different types of classification algorithms are adopted to use. This study specifically selects several DTs, NNs, and SVMs as individual classifiers.

C5.0 is one of popular decision-tree learning tools [28]. The distinctive characteristic of a C5.0 model is how the division rule split is chosen for the units belonging to a group, corresponding to a node of the tree [29]. The criterion is expressed as the following equation 4:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (4)$$

where *n* is the number of attributes, $|S_i|$ is the number of cases in the partition S_i, $|S|$ is the total number of cases in *S*. Entropy impurity is the usual choice as the following equation 5:

$$I(m) = -\sum_{i=1}^{k(m)} \pi_i \log \pi_i \quad (5)$$

where π_i is the fitted probabilities of the levels present in node *m*, which are at most *k(m)*.

Classification and regression tree (CART) partitions data into two subsets so that the records within each subset are more homogeneous than in the previous subset [13]. There are three different impurity measures used to find splits for CART models, including Gini, Twoing, and Least-squared deviation. For symbolic targets, Gini or twoing can be the choice. The Gini index, *g(t)*, at a node *t* in a CART tree is defined as:

$$g(t) = \sum_{j \neq i} p(j \mid t) p(i \mid t) \quad (6)$$

where *i* and *j* are categories of the target field. The Towing index, $\Phi(s,t)$, for a split *s* at a node *t* is defined as:

$$\Phi(s,t) = [\sum_{j} | p(j \mid t_L) - p(j \mid t_R) |]^2 \quad (7)$$

where t_L and t_R are the nodes generated by the split.

MLP is a hierarchical structure of several perceptrons and overcomes shortcomings of the single- layer neural networks [25][34]. The training algorithm for MLP requires differentiable, continuous nonlinear activation functions:

$$o = \frac{1}{(1+e^{-s})} \quad (8)$$

where *s* is the sum of products of weights and inputs.

A RBF neural network has an input layer, a hidden layer, and an output layer. Neurons in the hidden layer contain Gaussian transfer functions whose outputs are inversely proportional to the distance from the center of neurons. The Exhaustive prune method is related with a prune method. It starts with a large network and prunes the weakest units in the hidden and input layers as training proceeds.

SVM generates input-output mapping functions from a set of labeled training data. For classification, nonlinear kernel functions are often used to transform input data to a high-dimensional feature space. The RBF kernel, such as an exponential(Gaussian) kernel uses the Euclidean distance between vectors **x** and **y**:

$$K_{RB}(\mathbf{x}, \mathbf{y}) = k(\| \mathbf{x} - \mathbf{y} \|_2^2) = \exp(-\frac{1}{2\sigma^2} \| \mathbf{x} - \mathbf{y} \|_2^2) \quad (9)$$

where $k : R_0^+ \rightarrow R$ is a kernel profile. For a Polynomial kernel,

$$K_{Pol}(\mathbf{x}, \mathbf{y}) = (\gamma \mathbf{x} \cdot \mathbf{y} + r)^d, \gamma > 0 \quad (10)$$

The ensemble approach uses 'Simple voting' and 'Confidence-weighted voting'. With simple voting, if two out of three models predict 'yes', then 'yes' wins by a vote of 2 to 1. In the case of confidence-weighted voting, votes are weighted based on the confidence value for each prediction. Thus, if one model predicts 'no' with a higher confidence than the two 'yes' predictions combined, then 'no' wins.

IV. Data Analysis and Results

A. Data collection and preprocessing

Data involved in this study were based on the electronic medical records of chest pain patients who had received treatment in an emergency department of a hospital. Raw data of 478 patients were extracted from the database of the emergency department from July 2006 to June 2007. There were 410 kinds of lab tests. Because a patient sometimes took several tests at a time and/or took the same test more than once, therefore the total number of lab tests received by 478 patients corresponded to 16,581 records.

The EMR dataset also included eleven attributes available from the database of the emergency department, which were patient number (PNO), diagnosis, date of arrival, time of arrival, age, gender, symptom, ocode, subcode, oname, and testing results. To protect privacy of the patient, an artificial PNO was assigned uniquely to the individual patient. 'ocode' is a high-level classification code and 'subcode' is a low-level classification code that represents each lab test. To facilitate analysis, the data were sorted in ascending order of PNO by using EXCEL 2007 and rearranged to put the same patient's information in the same row.

Fig. 2 illustrates samples of preprocessed data with a row having the following attributes: PNO, diagnosis, arrival date and time, age, gender, symptom, and a series of 410 lab tests.

PNO	J503942 01	J25283A 14	J25283A 02	J25283A 03	J25283A 04	J25283A 05	J25283A 06	J25283A 07	J25283A 08	J25283A 09
2338023	0.050	16	7.800	1.800	200	25	0.600	40	3.500	1.900
2217640	0.040	28	8.100	2.000	175	5	0.700	100	5.600	3.000
2366723	0.060	5	7.500	7.500	172	117	3.400	104	5.300	2.900
2368974	$null$	19	8.800	3.500	142	16	1.000	170	7.100	4.000
1515254	0.040	4	7.600	3.700	86	25	1.100	180	5.700	3.200
636800	0.040	16	9.400	2.800	99	17	0.900	197	7.400	4.400
1667842	0.110	21	9.000	3.200	218	18	1.000	229	6.400	4.400
2330496	0.050	19	9.200	3.400	117	24	1.000	193	7.300	4.200
1442085	0.280	19	6.800	2.900	74	22	1.100	99	4.500	2.100
1339007	0.040	59	8.700	2.700	127	16	1.200	160	6.700	3.400
2072409	0.040	33	8.800	2.300	104	9	1.000	111	6.700	3.800
2265541	$null$	12	8.400	2.600	96	5	0.600	99	5.900	2.600
2357218	0.040	49	9.500	4.100	154	8	0.800	198	7.400	4.000
2353393	0.060	12	8.300	2.800	172	16	1.000	162	7.600	3.500
521271	0.050	12	9.900	3.700	142	13	0.900	173	7.800	4.400
1874000	0.040	38	8.800	2.800	136	13	0.900	125	5.800	4.000
2264552	0.040	13	9.400	4.300	185	31	1.500	129	7.700	4.200
2141978	0.070	19	8.800	2.700	113	9	0.800	215	6.600	4.000
1966197	0.040	16	9.000	2.700	119	21	1.400	183	7.200	4.000
2367199	0.040	15	8.900	4.100	66	12	0.700	172	6.800	3.900
2366907	0.090	12	9.400	3.900	142	13	0.900	149	7.400	4.400
2370167	0.850	15	8.200	3.600	113	11	0.800	189	6.200	4.100
2097356	0.060	29	10.000	3.400	117	23	1.100	135	7.200	4.400
2101946	0.040	12	9.300	3.900	100	14	1.000	130	7.900	4.400
2365670	0.120	53	9.100	2.900	120	17	0.900	184	5.800	4.300
2311563	0.040	30	9.000	4.100	199	13	0.900	168	7.000	4.100

Fig. 2 Preprocessed electronic medical records with lab test results

B. Description of data

The descriptive statistics of patient information are summarized as follows. The majority of patients were men. In terms of the patient age, more than 56.25% of patients exceeded 60 years. When chest pain diseases were clustered into three groups, such as AMI, AP, and others, the sample data showed that AMI was the most typical disease (45.82%). As previous studies have pointed out, the AMI disease has been a threat to human life with a higher risk [3].

There were 106 patients with the AP (22.18%) disease that has been the second critical disease and there were 153 patients with other chronic diseases (32.01%) that include Cardiomyopathy, Pneumothorax, Duodenal ulcer, and Respiratory tuberculosis, for example. Because the majority of patients were the old, this study focused on them in particular. Out of 478 patients, 180 patients who were older than 60 years old were selected for the further analysis.

C. Associative patterns among lab tests

As mentioned above, there were a series of 410 lab tests conducted in the focal emergency room. Not all patients, however, need to undergo all the lab tests since they arrive. The patient's condition could get worse while waiting to be tested. If a patient gets testing unnecessary for diagnosis, it could be a waste of time and money. In order to reduce the number of lab tests and to reduce time spent on tests, the most critical lab tests need to be revealed to diagnose the disease. In doing so, the domain knowledge in the field of medicine was combined to discover important lab tests.

According to Kenneth and Sharon [18], Butler and Swencki [3], and Ren [30], the AP and AMI diseases belong to the Acute Coronary Syndrome (ACS). Creatine Kinase (CK), Creatine Kinase MB fraction (CK-MB), and Troponin are very sensitive and critical tests to diagnose the ACS diseases. Based on their findings, association rules which contained each of CK (sub-code J252630_01), CK-MB (sub-code J252640_01), and Troponin (sub-code J503942_01) were discovered by the Apriori algorithm. In the process of analyzing the relationship between lab tests, sub-codes indicating specific lab tests were set as both input and target variables. Both minimum support and minimum confidence values were set at 0.9 to discover the rules showing a higher frequency.

There were 4,650 rules that met the threshold of support (0.9) and confidence (0.9). As a result of using the domain knowledge for the diagnosis of ACS, the number of rules which included three crucial lab tests (CK, CK-MB, and Troponin) has been reduced to 3,585 from 4,650. Table 1 shows samples of association rules at the sub-code level. Consider the relationship between lab tests J503942_01 and J252640_01, for example. The support value for the rule, which means 'If a patient takes a lab test J503942_01, then he also takes a test J252640_01,' was calculated as follows:

$$\text{Support}(J503942\_01 \rightarrow J252640\_01) = 325/327 = 0.9939$$

The confidence of the rule depended on its consequent and antecedent:

$$\text{Confidence}(J503942\_01 \rightarrow J252640\_01) = 325/326 = 0.9969$$

The lift was a normalized measure of interestingness,

$$\text{Lift}(J503942\_01 \rightarrow J252640\_01) = 0.9969/0.9969 = 1$$

TABLE I
SAMPLES OF CO-OCCURRENCE PATTERNS BETWEEN LAB TESTS AT THE SUB-CODE LEVEL

Confidence	Support	Lift	Association rules at the sub-code level
One antecedent → One descendent			
100	99.39	1	J503942_01 → J252640_01
100	99.39	1	J503942_01 → J252630_01
100	99.25	1	J503942_01 → J252620_01
Two antecedents → One descendent			
99.87	94.18	1.01	J252630_01 and J252640_01 → J25283A_13
99.87	94.18	1.01	J252630_01 and J252640_01 → J25283A_12
99.87	94.18	1.01	J252630_01 and J252640_01 → J25283A_11

From these association rules, it was found that 56 lab tests had a strong association with the critical lab tests, such as CK, CK-MB, and Troponin.

D. Clinical classification knowledge for chest pain diseases

For the building of diagnosis intelligence by using an ensemble approach, 180 patients' records were split into two kinds of data sets at random: a training dataset (126 records, 70%) and a validation dataset (54 records, 30%). Input variables included the selected 59 lab tests mentioned above and demographic information (i.e., gender, age) of each patient

International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:7, 2012

[16]. The target variable included three types of chest pain diseases: AP, AMI, and other chest pain diseases.

Table 2 summarizes the confusion matrix for the best final ensemble obtained on the training data set. Each row represented the actual diseases while each column represented the classified diseases in the confusion matrix. The diagonal values showed correct classifications.

TABLE II
CONFUSION MATRIX FOR THE BEST FINAL ENSEMBLE OBTAINED ON THE TRAINING DATA SET

		Classified diseases		
		Others	AP	AMI
Actual diseases	Others	42 (33.33%)	0	1 (0.79%)
	AP	0	41 (32.54%)	2 (1.59%)
	AMI	0	0	40(31.75%)

The classifications had an error rate of $100 \times (1 + 2)/126 = 2.38\%$. The classification accuracy was 100% (= 40/40) for the AMI disease, 95.35% (= 41/43) for the AP disease, and 97.67% (= 42/43) for other chest pain diseases. Total accuracy reached 97.62%.

V. VALIDATION OF MODELS AND COMPARISON

For the validation of the best final ensemble model, the records of 54 patients with chest pain were prepared separately from the training data set. Table 3 lists the classification accuracy of each single classifier and of each ensemble generated on the validation set of chest pain data.

TABLE III
CLASSIFICATION ACCURACY OF SINGE AND ENSEMBLE CLASSIFIERS (VALIDATION DATA)

Type of classifier	Aggregation method	Classifier	Classification accuracy (%)
Single		SVM (RBF)	87.50
		SVM (Polynomial)	83.93
		NN (MLP)	50.00
		NN (RBFN)	64.29
		NN(Exhaustive prune)	80.36
		CART (Gini)	76.29
		CART(Twoing)	80.36
		C5.0	83.93
Ensemble	Voting	SVM ensemble	85.71
		NN ensemble	76.57
		DT ensemble	85.71
	Confidence-weighted voting	SVM ensemble	82.14
		NN ensemble	80.36
		DT ensemble	83.93
	Voting	Final ensemble	92.59
	Confidence-weighted voting	Final ensemble	91.07

Among the single classifier models, the SVM (RBF) model showed the best performance on the validation data set (87.50%). For the voting ensemble, both SVM and DT ensembles had the best classification accuracy (85.71%). As in the case of training data set, the final ensemble model generated by voting exhibited the best accuracy (92.59%).

All models could be compared in terms of their gains. Gains are defined as the proportion of total hits that occurs in each decile. The larger the area between the gain chart and the diagonal line, the better the model under consideration. Fig. 3 shows the gain charts for single classifier models and the best final ensemble model being compared together (achieved on the validation data). The gain chart of the best final ensemble model suggests that it has better performance.

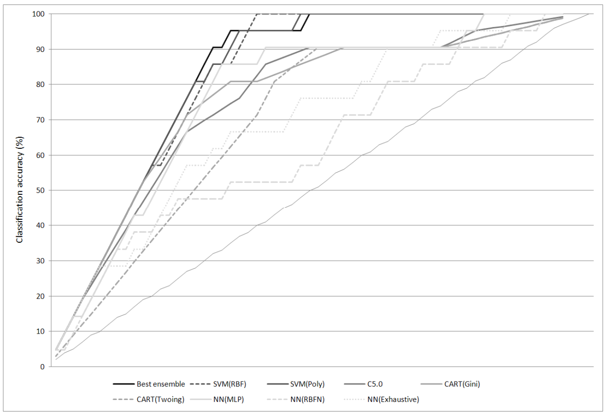

Fig. 3 Gain charts for the considered models obtained on the validation data

VI. CONCLUSIONS AND FUTURE RESEARCH

The purpose of this study was to build a hybrid model to derive diagnosis intelligence to aid in clinical decision making in a hospital environment. This study used an Apriori algorithm and an ensemble approach to combine different classification algorithms, and analyzed actual data collected from a hospital to classify chest pain diseases, which can help physicians to make clinical decisions faster and more accurately.

Through evaluation, this study showed that the final ensemble model performed well (92.59% of classification accuracy) in diagnosis of chest pain diseases. In order to show the advantage of combining the association rule mining and the ensemble model, a further analysis was conducted by using the three critical lab tests only as input variables. The classification accuracy of using 59 lab tests as input variables had better accuracy. By integrating medical domain knowledge into the research, unnecessary lab tests could be filtered out so that this has brought higher accuracy to the classification of diseases and savings on time and money spent on the processing of lab tests.

Future research should consider several improvements. First, this study adopted a hybrid decision-model approach, combining the association rule mining and the ensemble of classifiers. However, another combination of data mining techniques is still possible to accomplish the same task of medical data mining. To think of an alternative methodology will be a good challenge. Second, as the accumulation and diversification of medical data increase, the need for adaptive methods of medical decision-making has been gradually expanding. Considering the current situation of hospital information systems, more complete and comprehensive systems will be necessary for the future.

Such a system should have medical intelligence mentioned in this study, which can be enhanced and expanded. For example, it can incorporate other medical attributes such as image or audio, which can help to raise the accuracy of classification and prediction of diseases. Third, the data used in this study were collected from a single hospital in a city during one year. Due to the geographical and temporal limitation, the data may not be typical for all chest pain patients. Even though the distribution of the sample data obtained was consistent with that of patients with chest pain in the country, those efforts are needed to generalize the classification model and methodology.

REFERENCES

[1] Abdullah U, Ahmad J, Ahmed A (2008) Analysis of effectiveness of apriori algorithm in medical billing data mining, in Proceedings of 4th International Conference on Emerging Technologies.
[2] Bassan R, Scofano M, Gamarski R, Dohmann HF, Pimenta L, Volschan A, Araujo M, Clare C, Fabricio M, Sanmartin CH, Mohallem K, Gaspar S, Macaciel R (2000) Chest pain in the emergency room: Importance of a systematic approach. Arquivos Brasileiros de Cardiologia 74(1):13-29.
[3] Butler KH, Swencki SA (2006) Chest pain: a clinical assessment. Radiologic Clinics of North America 44(2):165-179.
[4] Ceglowski R, Churilov L, Wasserthiel J (2007) Combining data mining and discrete event simulation for a value-added view of a hospital emergency department. Journal of the Operational Research Society 58:246-254.
[5] Cios KJ, Moore GW (2002) Uniqueness of medical data mining. Artificial intelligence in medicine 26(1-2):1-24.
[6] Conforti D, Guido R (2005) Kernel-based support vector machine classifiers for early detection of myocardial infarction. Optimization Methods and Software 20(2-3):401-413.
[7] Conroy RM et al. (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European Heart Journal 24(11):987-1003.
[8] Duguay C, Chetouane F (2007) Modeling and improving emergency department systems using discrete event simulation. Simulation 834):311-320.
[9] Ellenius J, Groth T (2000) Transferability of neural network-based decision support algorithms for early assessment of chest-pain patients. International Journal of Medical Informatics 60(1):1-20.
[10] Erhardt L, Herlitz J, Bossaert L, Halinen M, Keltari M, Koster R, Marcassa C, Quinn T, van Weert H (2002) Task force on the management of chest pain. European Heat Journal 23:1153-1176.
[11] Fayyad UM (1997) Data mining and knowledge discovery in databases: implications for scientific databases, in Proceedings of Ninth International Conference on Scientific and Statistical Database Management.
[12] Fromm RE, Gibbs LR, McCallum WG, Niziol C, Babcock JC, Gueler AC, Levine RL (1993) Critical care in the emergency department: a time-based study. Critical Care Medicine 21(7):970-976.
[13] Giudici P (2003) Applied Data Mining: Statistical Methods for Business and Industry. John Wiley & Sons, Hoboken.
[14] Guven A, Kara S (2006) Classification of electro-oculogram signals using artificial neural network. Expert Systems with Applications 31(1):199-205.
[15] Jegelevicius D, Lukosevicius A, Paunksnis A, Barzdziukas V (2002) Application of data mining technique for diagnosis of posterior Uveal Melanoma. Informatica 13(4):455-464.
[16] Kannel WB, Vasan RS (2009) Is age really a non-modifiable cardiovascular risk factor. American Journal of Cardiology 104(9):1307-1310.
[17] Kantardzic MM, Zurada J (2005) Next Generation of Data-Mining Applications. John Wiley & Sons, Hoboken.
[18] Kenneth HB, Sharon AS (2006) Chest pain: A clinical assessment. Radiologic Clinics of North America 44:165-179.
[19] Khan FS, Anwer RM, Torgersson O, Falkman G (2008) Data mining in oral medicine using decision trees. World Academy of Science, Engineering and Technology 37:225-230.
[20] Kononenko I (2001) Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine 23(1):89-109.
[21] Lin WT, Wang ST, Chiang TC, Shi YX, Chen WY, Chen HM (2010) Abnormal diagnosis of emergency department triage explored with data mining technology: An emergency department at a medical center in Taiwan taken as an example. Expert Systems with Applications 37(4):2733-2741.
[22] Majumder SK, Ghosh N, Gupta PK (2005) Support vector machine for optical diagnosis of cancer. Journal of Biomedical Optics 10(2):24-34.
[23] Martinez-Selles M, Bueno H, Sacristán A, Estévez A, Ortiz J, Gallego L, Fernández-Avilés F (2008) Chest pain in the emergency department: incidence, clinical characteristics and risk stratification. Revista Espanola de Cardiologia 61(9):953-959.
[24] Masuda G, Sakamoto N, Yamamoto R (2002) A framework for dynamic evidence based medicine using data mining, in Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems.
[25] Mitchell TM (1999) Machine Learning and Data Mining. Communications of the ACM 42(11):30-36.
[26] Motoda H, Ohara K (2009) Apriori. in: V. Kumar (Eds.), The Top Ten Algorithms in Data Mining. Chapman & Hall/CRC, FL, 61-92.
[27] Qiao DD (2009) Clinical diagnosis of life threatening chest diseases: based on 156 cases. Journal of Chinese Modern Medicine.
[28] Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers.
[29] Ramakrishnan N (2009) C4.5. in: V. Kumar (Eds.), The Top Ten Algorithms in Data Mining. Chapman & Hall/CRC, FL, 1-19.
[30] Ren H (2007) Clinical diagnosis of chest pain. Chinese Journal for Clinicians 36:5-7.
[31] Riccardo B, Blaz Z (2008) Predictive data mining in clinical medicine: current issues and guidelines. International Journal of Medical Informatics 77:81-97.
[32] Ridker PM, Hennekens CH, Buring JE, Rifai N (2000) C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. New England Journal of Medicine 342:836-843.
[33] Rossille D, Cuggia M, Arnault A, Bouget J, Le Beux P (2008) Managing an emergency department by analyzing HIS medical data: a focus on elderly patient clinical pathways. Health Care Management Science 11(2):139-146.
[34] Shmueli G, Patel NR, Bruce PC (2007) Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner. John Wiley & Sons, Hoboken.
[35] Statistics Korea (2010), http://www.kostat.go.kr/.
[36] Tan Y, Yin GF, Li GB, Chen JY (2007) Mining compatibility rules from irregular Chinese traditional medicine database by Apriori algorithm. Journal of Southwest JiaoTong University 15(4).
[37] Velagaleti RS, Massaro J, Vasan RS, Robins SJ, Kannel WB, Levy D (2009) Relations of lipid concentrations to heart failure incidence: the Framingham Heart Study. Circulation 120(23):2345-2351.
[38] Yun YP (2008) Application and research of data mining based on C4.5 Algorithm. Harbin University of Science and Technology.
[39] Zaki MJ (2004) Mining non-redundant association rules. Data Mining and Knowledge Discovery 9(3):223-248.