A Hybrid Mining Approach to Facilitate Health Insurance Decision: Case Study of Non-Traditional Data Mining Applications in Taiwan NHI Databases

Joseph Tan,
Professor of eHealth, McMaster University,
tanjosep@mcmaster.ca

Fuchung Wang,
Senior Executive Officer, Taipei Division, NHIA
jasonwfz@nhi.gov.tw

Abstract

This study examines time-sensitive applications of data mining methods to facilitate claims review processing and provide policy information for insurance decision-making vis-à-vis the Taiwan National Health Insurance databases. In order to obtain the best payment management, a hybrid mining approach, which has been grounded on the extant knowledge of data mining projects and health insurance domain knowledge, is proposed. Through the integration of data warehousing, online analytical processing, data mining techniques and traditional data analysis in the healthcare field, an easy-to-use decision support platform, which will facilitate the health insurance decision-making process, is built. Drawing from lessons learned in case study, results showed that not only is hybrid mining approach a reliable, powerful, and user-friendly platform for diversified payment decision support, but that it also has great relevance for the practice and acceptance of evidence-based medicine. Researchers should develop hybrid mining approach combined with their own application systems in the future.

1. Introduction

Fueled by massive amount of data being collected on a daily basis, the resulting information deluge faced by healthcare insurance systems characterized by a single payer (monopolistic) and/or a few payers (oligopolistic) has, in turn, caused traditional information retrieval methods and data analysis to perform inadequately. As a result, a new interdisciplinary field of data mining approaches, encompassing both classical statistical methods and modern machine learning tools to support efficient and effective mining of information and the discovery of knowledge (and hidden patterns) from databases [1][2] is gradually being realized and implemented. Using health data analytics and data mining techniques to make business decisions for increasingly complex healthcare organizations and

systems can effectively influence cost, revenues, and operational efficiency while maintaining a high level of care.

To date, various data mining methods and application systems approaches have been promoted. Unlike the traditional statistical and online analytical processing (OLAP) tools, these newer techniques have integrated algorithmic and inductive approaches to facilitate advanced data analysis and decision support to yield business intelligence [3]. Some researchers have explored the use of data mining in the development of a decision support system to manage health care services [4]. The others have provided the outcome of mining methods in the value of quality improvement [5]. In 2014, the US Federal government spent 1.1 trillion for Medicare and Medicaid. By 2030, expenditures for these two programs are projected to consume 50% federal budget. Cost and spending factors drive need for change. Given that health care organizations are typically characterized as complex systems with multi-stakeholders, they have also known to be highly information dependent; therefore, knowledge intensive technology is vital to the survival of these complex organizations and systems.

Hybrid mining approach [6] underlies a novel design for uncovering knowledge embedded in complex data systems linking multiple stakeholders as exemplified by the National Health Insurance (NHI) databases in Taiwan. The National health expenditures in Taiwan was 6.6% of GDP in 2012, simultaneously United States was 16.2%. Instead of higher expenditures, life expectancy of US is still lower than Taiwan at least 2 years. The NHI is a monopolistic social insurance plan, a mandatory healthcare plan characterized by a centralized disbursement of funds and its administration cost is only 1% of total expenditures, financed primarily through direct government funding, employment premiums and user co-payments. In 2002, in light of the need to avoid ongoing losses and for cost containment, NHI moved to alter the payment system from a fee-for-service to a global budget, foreshadowing a kind of prospective payment system. Meanwhile, Taiwan had plans to reform its payment schema to include case payment, pay-for-performance and diagnostic related group (DRG) under the global budget [7]. All claims

H♥CSS

are filed and processed electronically. The NHI Administration's automated IT-supported claims review checks for the overall appropriateness of claims and also selected a small percentage of all claims for individual professional review by clinical experts [8].

Even so, it is important to clarify the underlying complex relations among embedded entities in the NHI databases in order to help decrease the nature of obstacles in the healthcare application. Accordingly, the goal of this paper is to showcase the hybrid mining approach based on the extant knowledge of data mining projects and health insurance domain knowledge. Specifically, our research validates the approach by drawing on lessons learned in case study facilitating evidence-based medicine to support multiple health insurance payment strategies.

2. Hybrid Mining: An Overview

Briefly, data mining may be defined as the iterative process of analyzing data via either automatic or manual methods in order to discover implicit, but useful information and previously unknown patterns and relationships hidden in data. Today, many organizations are using data analytics and data mining to improve on product and/or service marketing capabilities, to detect abnormal transactional patterns, and to predict future directions based on past experiences and current trends. In adapting to rapidly changing environments, new and rapidly expanding organizations are clearly in need of this technology.

As real-world problems tend to be highly nonlinear in nature, it is difficult, if not impossible, to develop a comprehensive model taking into account all of the independent variables impacting various organizational decisions using only conventional statistical approaches; instead, complex data analytics and new forms of data mining approaches have been demonstrated to be more reliable and useful [9]. Meanwhile, the digital revolution has provided relatively inexpensive and available means to collect and store large amount of current and historical data. As databases grow in size, it will be increasingly more difficult to support all organizational activities, even routine decision making. Hybrid mining will be an efficient way and has been proposed for health insurance database [10].

As traditional manual data analysis has its limitations, new methods for advanced data analysis and knowledge discovery have become indispensable, thereby promoting a new field of data mining. These new approaches are based on algorithms integrating complex statistical tools, artificial intelligence, neural networks, inductive rules, and predicate logic. In contrast to the traditional reactive decision support system tools, these newer approaches will initiate analysis to create knowledge, as the premise underlying data mining is proactive. Broadly speaking, therefore, data mining or

data analytics can be applied to solve specific problems that have not yet been identified by the user.

2.1. Hybrid Mining Approach

The promise of hybrid mining approach as the next generation of data mining systems has already been discussed in the extant literature. The approach is based on a novel combination of existing approaches, as noted by a number of researchers [11]. This work is motivated by the obvious limitations of classical paradigms for data mining that have been used in isolation. As each method carries its own advantages and shortcomings, mixing the use of different methods can be both complementary and desirable. Hybrid approach shows improved data analytic modeling accuracy and robustness [12]. Research into the integration of different methods and case applications of hybrid mining technologies will continue to broaden our knowledge of this evolving field.

Applying clustering to refine the design of neural networks is a concept that originated in the mid-1990s [14][15]. Essentially, hybrid design architecture use neural networks (NNs) guided by clustering algorithms to generate more accurate classifiers. Following the success of the hybrid mining application, there has been increasing number of studies investigating hybrid mining approaches. Examples of tasks that can be facilitated by the use of the hybrid mining include constructing the credit scoring models from a credit database and designing an effective credit-scoring model. West [15], for example, investigated the credit scoring accuracy of five NN models, and reported that non-parametric and hybrid design architecture would be very useful in developing effective credit scoring models. He then concluded that hybrid architecture should be considered for credit scoring applicants.

Kuo and his colleagues [16] developed a two-stage method by first allowing self-organizing map to determine the number of clusters and then, employing the K-means algorithm to classify samples into cluster. As well, Lee, Chiu, Lu, and Chen [17] integrated the back-propagation NNs with traditional discriminant analysis approach to achieve greater accuracy and easier interpretation of mining results. Chen and Huang [18] also presented a work involving two interesting credit analysis problems and resolved them by applying NNs and genetic algorithms techniques. Hsieh [19] proposed a hybrid system based on clustering and NN techniques, leading to a hybrid architecture that generates more accurate classifiers. Lee and Chen [20] proposed a two-stage hybrid modeling procedure with artificial NNs and multivariate adaptive regression while Huang and Wang [21] combined genetic algorithms with support vector machine classifier. The results of these research showed that hybrid mining is a promising addition to the existing repertoire of data mining methodology.

Hybrid mining approach is well developed not only in technological but also in applied studies. For example, it has been a great progress in the application of data mining techniques focus on credit scoring competition of the consumer credit market [15][17][28][20]. Sadatrasoul and his colleagues [22] investigated hybrid methods focus on credit scoring from 2000 to 2012 in four categories and found two of the frequently used combinations are "classification and classification" and "clustering and classification". These hybrid methods can identify and extracts potential good and bad applicants groups and the scoring model also can apply to detect patient visits with abusive utilization pattern based on profiling information extracted from insurance claims database.

2.2. Enhanced Health Care with Better Mining

The routine delivery of health care services generate massive streams of both administrative and clinical data about patients, hospital procedures, bed costs, and claims processes. In this age of rapid knowledge profusion and digital transformation, and given that data, in and of themselves, represent strategic resources for health care institutions, the use of data analytics and data mining in many areas of health care services delivery to track patient treatment effectiveness, improve the health care procedures and management, enhance customer relationship management and detect frauds and abuses [23], is something that cannot be overly emphasized.

As clinicians and managers continue to find values in data analytics and new data mining approaches through their respective power to unlock new patterns [23][24] in data, leading to potential improvement of organizational resource utilization and patient health, health care organizations are moving to implement data mining technologies to help control costs and improve the efficacy of patient care. Moreover, as these newly discovered patterns are oftentimes based upon evolving clinical practices, they can, in turn, be used to set new standards of evidence-based care. Broadly speaking, health data analytics and data mining offer community-based gains that improve healthcare forecasting, analyses, and visualization [25].

If healthcare decision makers can make clever use of health data analytics and data mining techniques, new knowledge can be automatically generated on a timely basis. For example, a patient's length of stay (LOS) [26] is a major component in the costing of inpatient treatment. Based on the premise that it is possible to reduce costs through reduced inpatient LOS, the knowledge acquired will, in turn, lead automatically to a reduction of cost. Another example is applying in the medical prescription. Although the health professionals prescribed for treating diseases are diversity, mining of those data from a nationwide prescription database will be a great benefit to develop a smart medication

recommendation model to reduce costs and improve patient safety [27].

Put together, a novel hybrid data mining approach for knowledge extraction and classification in health services and medical databases will benefit the coordination of health services delivery and health care management significantly. The approach might combine self-organizing map, K-means with NN-based classifier to clustering all data and fusing them through the use of serial and parallel fusion in conjunction with neural classifier [28]. Besides, association rule is also a popular method to add into hybrid mining approach. Valuable knowledge can be discovered from application of data mining techniques in healthcare database. These new applications can be naturally classified into three broad categories, include of descriptive, predictive and prescriptive model, in order to increase the quality of care provided to the patients and decrease the medical expenditure.

3. Hybrid Mining to Facilitate Health Insurance Payment

With escalating healthcare costs over the past years, timely analysis of healthcare information has become an issue of growing importance. Addressing the effectiveness of association rules and neural segmentation in analyzing and retrieving unknown behavioral patterns from huge datasets, the Australian Health Insurance Commission (HIC) had collected detailed claims information and established a homogeneous claim database through the administration of various programs. As the medical reimbursement process was time consuming, this specific case had shown that health data analytics and advancing data mining algorithms could be used successfully on large, real-time customer data, with reasonable execution time. Moreover, the applied HIC project had also demonstrated that the intelligent applications of these algorithms could result in quantifiable benefits for the interested and participating organizations, thereby helping to identify specific actions that have to be taken [29].

3.1. Applying data mining to payment information system

Advancing health data analytics and data mining algorithms may be used to predict health outcomes and provide policy information. Chae and his colleagues [30] demonstrated how decision tree algorithm could be used in a policy analysis for hypertension management and the association rules that provided an occurrence relationship among risk factors. While logistic regression provides insights into general risk factors for hypertension, it does not provide specific information on the segment characteristics of age or risk factors that may be useful for policy analysis. Moreover, the

comparison of the how different data mining methods performed was not consistent.

Another research explored the use of artificial NNs in the development of a decision support system to manage healthcare non-clinical services. The advantage of using this system is that healthcare managers can provide their own risk assessment values such as point score system based on their own knowledge and corporate objectives [3]. While the development and application of decision models that is based on a data mining method, applying the data mining to decision support through supporting structure and utility function development from pre-classified patient's data may lead to the discovery of hierarchical decision models and improved performance both in terms of classification accuracy and the discovery of meaningful concept hierarchies [31].

Simply stated, decision making is a process of determining the choice. "Decision support" means using the methods and tools for supporting people (decision makers) involved in the decision-making process. Data mining and decision support approaches (including data analytics) are often complementary and supplementary. In this sense, both the information from retrospective data and existing medical knowledge are combined in data analytics and data mining to construct potentially useful and highly predictive medical decision model.

Claims-based databases or health care utilization databases are made of routinely collected records originally intended for management purposes. They are the main component of healthcare payment information system for reimbursement of health insurance. In recent decades, some national data holders of administrative health care databases started to recompile these databases and made these available for academic research purposes. One of the most well-known administrative health care databases is National Health Insurance Research Database (NHIRD) in Taiwan. NHIRD studies were also published diversely in a growing number of journals [32]. However, most of them used traditional data analysis to describe or predict health outcomes and the relationships and only a few applied data mining methods for complex computation. Inappropriate payments by insurance organizations or payers occur because of errors, abuse and fraud. The scale of this problem is large enough to make it a priority issue for health systems. Joudaki and his colleagues recommended general steps to data mining of health care claims [33]. Shin and his colleagues [34] proposed the scoring model to detect outpatient clinics with abusive utilization patterns based on profiling information extracted from insurance claims database. Lin and Yeh [35] take advantage of data-mining technology to design models and find out cases requiring for manual inspection so as to save time and manpower.

Health care insurance programs rely on the database to develop knowledge discovery processes not only for fraud and abuse detection but also for payment policies. However, traditional methods of detecting health care fraud and abuse are time-consuming and inefficient [33], and moreover, they are not enough to support different payment decisions in the different decision level. Hybrid mining approaches use real claims-based databases will have an adequate statistical power due to their large size. Meanwhile, calculation of incidence rate is possible because of long follow-up period. If the databases are nationwide, population-based design can help to have little chance for selection bias.

3.2. A Framework for Payment Decision Support

Achieving a framework for health insurance payment decision support will serve to crystallize the role of hybrid mining based systems in health insurance payment. Based on a review of the extant literature and to ensure that the design and implementation of support systems be available in practice, our research applied the design perspective to synthesize and propose such a framework. Two of the frequently used mining technologies clustering and association rule add into hybrid mining approach which identifies the most cost effective payment decision form scoring model.

This framework, as depicted in **Figure 1,** was based on the extant knowledge of hybrid data mining and health insurance payment projects. The framework combined data mining and decision support to be used as a template for quantifiable and/or subjective evaluation to be made by researchers and practitioners. As shown, the original method to process the health insurance payment decision follows the sequence of dotted lines in below which was automated or semiautomatic claims review checks for selected a small percentage of all claims for professional review by clinical experts. The framework integrated with hybrid mining to improve the efficiency of payment decision in the review process.
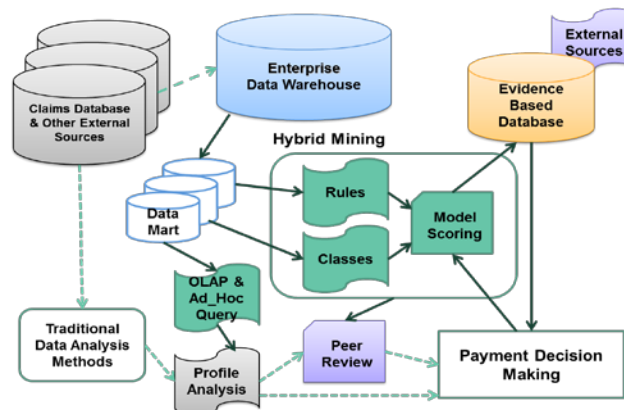
**Figure 1. The framework for health insurance payment decision support**

4. NHI case study

The National Health Insurance (NHI) system of Taiwan was launched in 1995 and established to provide healthcare insurance for all residents living in Taiwan. NHI is a compulsory program; therefore, 99.9% of the population is enrolled, including the nationals and foreign workers and even included inmates since 2013 [8]. In the health delivery side, about 93.5% of the providers have joined the NHI program. The NHI provides a uniform package for all insured, and the benefits are very comprehensive, covering a wide range of services even though Taiwan spends only 6% of its GDP on health care. Patients are free to choose providers and there is no significant waiting time to see a doctor. Meanwhile, medical expenditures are contained by the global budget payment system [7][8]. As a single-payer system, the National Health Insurance Administration (NHIA) is the only agency authorized to administer the program and collects premiums from the insured, their employers, and the government based on individual's monthly incomes.

Because the NHI is a single payer system, the administration cost is relatively low. It is less than 1% of the NHI medical expenditures, which is low compared to many Western countries. A key factor leading to the program's success is the adaptation of information system from its inception [7][8]. With its modernized health IT infrastructure and platform, it should be easier for Taiwan than it is in the US to enhance the cost effectiveness and the quality of health care [36]. This is because not only is 99% of the rate of claims processed through electronic media since 1996, but the smart-card program to connect multi-providers in health care has also been successfully implemented by 2004. Given the state of health data digitization, the NHIA can track almost in real-time what goes on in the nation's health care system.

The primary NHI payment information system (PIS) structure is shown in **Figure 2**. Different from some other countries, patients don't need to pay directly except out-of-pocket when they visit. The health care providers would send the claims to NHIA for payment in the reimbursement process. It will be benefit for patient because reduces their financial burden when they visit but increases the pressure of the cash flow form providers side. As the NHIA reimburses medical payments to healthcare providers on a monthly basis, the PIS is critical as the IT operating environment changes dynamically and grows increasingly competitive. The reimbursement system discussed next has been designed to facilitate the efficiency and cost-effectiveness of Taiwan's claims review process. Furthermore, the whole system is subject to strict privacy and security controls through the NHI Virtual Private Network that only NHI-contracted health service providers can use. The PIS are looking forward to support NHI program to have continuously succeeded in providing high-quality medical services with the comparatively low health expenditure.
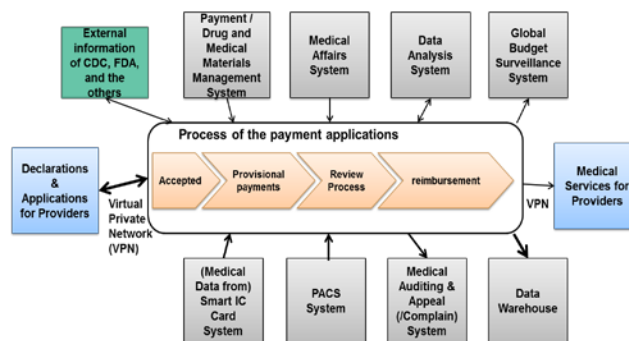
Figure 2. The NHI Payment Information System Structure

In 2015, there were more than 1 million outpatient visits and almost 90 thousand inpatients daily. The number of claims for outpatient treatment and medical orders was 43 million and 160 million per month respectively and the number of claims for outpatient was 300 thousand per month. Those claims need to be reimbursed within 2 months. The monthly provisional payment was NT$ 50 billion (US dollars was about $1.6 billion). In order to manage the utilization of medical resources and to assure the care quality, the NHIA develops various review mechanisms to examine the appropriateness of the submitted medical claims [37] and complete the procedural reviews.

Apart from prospective and on-site reviews, retrospective review is the key mechanism used by the NHIA. With retrospective review, two approaches are adopted, namely, procedural review and professional review. The procedural review is fully automated that includes auto-adjudication and profile analysis. In this process, NHIA check the accuracy and completeness of claim data electronically. NHIA already developed more than one thousand rules of auto-adjudication and hundreds indicators for profile analysis. In professional review, an estimated total of 3,000 physicians with various specialties are contracted with NHIA to peer review the rationality of medical procedures and services. To date, the deny rate of primary review stands only around 0.28%~2.66% in 2014. Taiwanese rely heavily on the government-managed health insurance system to share a large part of their healthcare expenses. With growing demands on NHIA payments, however, the fraudulent and abusive claim behaviors of some health services providers have challenged the integrity of the system.

A process data mining framework that utilizes the concept of clinical pathways to facilitate automatic and systematic construction of an adaptable and extensible detection model, which have been evaluated objectively by a real data set gathered from the NHI program,

provides the critically needed system safeguard. Indeed, past empirical experiments have demonstrated that this detection model is efficient and capable of identifying some fraudulent and abusive cases that were not detected via a manually constructed detection [38].
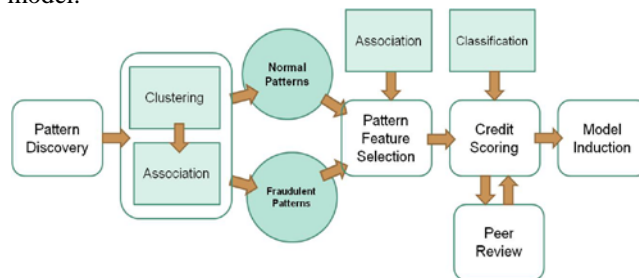
4.1.Decision support for the reimbursement process

After the NHIA was founded, it established the PIS to review the delivery of medical services to prevent waste and safeguard quality. Information Technology is used in conducting both procedural review and professional review, but especially the procedural review, which relies on profile analysis based on specific medical criteria to conduct automated audits, improving the efficiency of claim reviews. The traditional data mining were used occasionally but it is difficult to implement into the reimbursement process. For example, the mining association rules are popular and well researched methods for discovering relation between variables in NHIRD [39], but there remained a lack of applicable standards for evaluating highly specialized medical services. Most of mining literatures focused more on the technical methods and paid little attention to the practical implications of their findings for decision makers [33].

The decision support provided by the application of hybrid mining to facilitate claims review and payment processing will be more specialized and precise. Hybrid mining can help NHIA to extract useful information from millions of claims and identify a smaller subset of the claims or claimants for further assessment. The objective of claims review under global budget system has shifted gradually from the containment of medical expenditure to the assurance of medical care quality. To assure and improve on the care quality, the NHIA has to maintain an ongoing, strong functioning profile analysis database with data mining methods as well as a management model for case anomalies.

The healthcare quality indicators have also been identified and the decision support system has also been designed to analyze and monitor trends of quality indicators with tutorial for quality improvement activities embedded in the system. In addition, the NHIA will continue to entrust peer review to the medical associations to ascertain joint management with medical providers. Furthermore, the NHIA will keep adding more quality indicators to monitor the performance of health care providers and to disclose all these quality information to the public so as to reduce the asymmetry of medical information between health care providers and the general public. A process data mining framework is shown in **Figure 3**, which is combine the numerous monthly claims into a provider profile claims. According to the past claims data of each provider, the process could identify the group of different provider. For example, the high fraud

suspicious group should take more resources on the professional review. Meanwhile, the review results also provide a chance to verify and modify the credit scoring model.

Figure 3. The hybrid mining based decision support for reimbursement process

To fit in with NHIA newest strategy, a hybrid mining approach was considered as a solution to reduce the problems and increase the effectiveness when using the mining techniques. In particular, the aim was to create a better application mimicking the real world. This case presents a hybrid mining approach in mimicking physicians' behavior for healthcare management. The design was based primarily on clustering and NN techniques. Effective segmentation of GPs' practice patterns makes it easier to detect and investigate health fraud by recognizing and quantifying the features of claims and providers.

We used clustering techniques to preprocess the input data and used NNs to construct the scoring model. The clustering stage involved a classification process. A self-organizing map clustering algorithm was used to automatically determine the number of clusters and the starting points of each cluster. Then, the K-means clustering algorithm was used to generate clusters of samples belonging to new classes. In the NN stage, samples with new class labels were used in the design of the scoring model. Compared with the traditional mining approach on the basis of real physicians' behavior data sets, the proposed method demonstrates that the hybrid mining can be used to build an effective scoring model to develop a data-driven initiative to analyze and model the formal relations between GP' practice patterns and health fraud.

Table 1. A sample of scoring result

Credit scoring Cluster(3) 710 Physicians(16.4%) High variation in diseases High price per claim		
Fraud possible	Cluster(3-1) 16 belongs to Class A Providers	Cluster(3-2) 45 belongs to Class B providers
	Cluster(3-1) 420 belongs to Class A providers	Cluster(3-2) 229 belongs to Class B providers
Inappropriate possible	Cluster(3-1) 90 belongs to Class A Providers	Cluster(3-2) 125 belongs to Class B providers
	Cluster(3-1) 346 belongs to Class A providers	Cluster(3-2) 149 belongs to Class B providers

A sample of scoring result is shown in **Table 1**; we consolidated these ideas and implemented them into the reimbursement process with the decision support function. Following structured interviews with NHIA officers who were responsible for retrospective review process, we learned that the hybrid model was perceived as superior to the single mining method as provider performance assessment needs to take into account patients characteristic, provider characteristic, correlation between outcomes of patient within the same provider and number of patients in the provider. These data characteristics motivate the centrality of hybrid mining analysis. Moreover, the hybrid model applies data mining techniques to support the key policymakers in Taiwan government in managing and building up the healthcare management knowledge from submitted claims information of healthcare providers. The scoring results also feedback to providers via VPN, help them to know their situation. Apparently, not only did the hybrid model proved to contribute new knowledge, but the result of applying the knowledge as a means of changing physicians' behavior would serve as a good reference for applications to other national single-payer healthcare systems which most of outliers would be back to normal according to the following reviews

4.2. Decision support for quality of pay for performance

Over the years, NHIA had engaged in a number of quality improvement initiatives. Among five of the pilot programs implemented in the late 2001, the pay for performance (P4P) program was relatively popular. Essentially, the program permitted voluntary participation by primary care physicians, hospitals, or clinics and other care facilities. To participate in the program, providers must adopt some quality assurance steps such as meeting the qualification or certification requirements, following the treatment guidelines and certain other protocols. Apart from sending in the claims data, these providers must also store their self-reported performance data electronically in a website that is linked directly to NHIA for easy access and use. The NHIA will then review the claims with other fee for services claims monthly and finish the reimbursement within 2 months. And the self-reported performance data based on the outcome or process depending on how the program quality requirements have been satisfied, these reviews will then result next year in whether the claims bonus payment would be reimbursed or not.

An example of P4P is Diabetic care, which falls into a process-based payment category. Taiwan's single-payer system is an ideal platform for P4P program as the system embodies an information infrastructure that yields comprehensive and up-to-date information on the actual care delivered to patients. Although some

researches already applies data mining techniques to detect fraudulent or abusive reporting by healthcare providers using their invoices for diabetic outpatient services. However, health care quality improvement is the goal of these reimbursement plans.

Specifically, in our case, we collected the Diabetic care cases claims and self-reported performance data together to make a classification. The diabetic data in this instance was those captured for the northern area of Taiwan with 6,274 diabetic patients. The diabetic registries and databases systematically collected patient information, while our analysis showcased a method of applying data mining techniques, and some of the data issues, analysis problems, and results.

Table 2 presents the optimal clustering result in our specific case. Only the most important variables are depicted in the table. A total of five clusters that were found to be statistically significant are shown. Our next step was to join the historical claims data with the five clusters to identify the proper payment with multilevel analysis which one is patient level within provider and another is between providers. Following the different payment analysis, the results and methods would subsequently yield a solution for the policymaking not only in Diabetes but also for other diseases management.

Table 2. Clustering result in Diabetic patients and how the result apply to policy making

CLUSTER	Patients	Gender (F)	Family History	Tri glycerides	HbA1c	Pathological Changes of Foot	Standard Payment	Payment Range
A	1,899	100%	42%	180.76	8.529	84%	60	±10
B	964	1.04%	100%	188.9	8.271	71%	45	±6
C	1,337	0.67%	49%	198.41	8.65	100%	50	±6
D	1,280	76.88%	50%	188.66	8.61	87%	35	±3
E	794	55.09%	93%	246.75	9.642	0%	55	±6

The example discussed above is very similar to the research that has applied data mining method to classify consumers' behavior in choosing among Taiwan hospitals but take into account patient risk adjustment will have a normative commitment by the providers. Here, factors of consumers' behavior were sorted into four categories. Then, a back propagation NN classification model was developed. The model demonstrated the usefulness of 85.1% classification rate in sorting out consumers' styles. Based on the research findings, the evidence was enough to suggest that the NN model is useful in identifying existing patterns of hospitals' consumers.

4.3. Decision support for implementing the DRG policy

In contrast to above cases, outlines decision support benefitting the implementation of diagnostic related group (DRG) policy. Traditionally, past applications of

the data mining methodology have often to do with the analysis of one DRG by OLAP [4]. In the case reported here, we will introduce an integrated method via hybrid mining approach to aid the DRG policymaking.
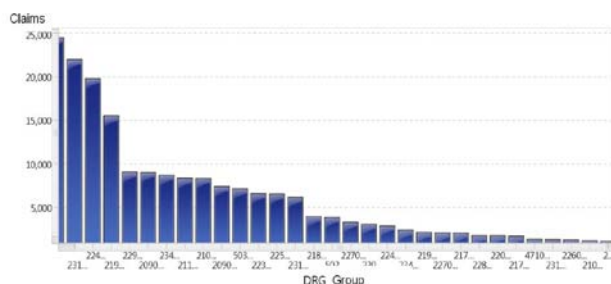
Figure 4. The claim of DRG for orthopedics

DRGs are divided into 25 major diagnostic categories(MDC) and relative weight (RW) is a value assigned to each DRG to reflect relative resource consumption for the group. The RW is used to calculate the total payment for the case. As part of expensive medical materials paid by patients will not count to DRG, some hospitals would like to abandon those claims. It will impact the RW calculation and the payment decision will have bias in the long term. A orthopedics DRG sample of hybrid mining approach combining automated methods and statistical knowledge is shown in **Figure 4** and **Table 3** Data mining is a core of the payment decision support process. Hybrid mining approach can help NHIA to extract useful information from thousands of claims and identify a smaller subset of the claims or claimants for further assessment.

Table 3. Comparison between hybrid mining model and the hospital claims

Hospital	Claims	Model Consistent		Percentage of Low Consistent
		<50%	>50%	
A	41	9	32	22.0%
B	43	9	34	20.9%
C	129	18	111	14.0%
D	111	12	99	10.8%
E	123	18	105	14.6%
F	26	3	23	11.5%
G	16	5	11	31.3%
H	4	1	3	25.0%
I	17	2	15	11.8%
J	10	1	9	10.0%

5. Conclusion and discussion

5.1. Management research and practice implications

Altogether, our research supports the idea that hybrid mining based system can serve as an efficient decision support in health insurance reimbursement decisions. The hybrid mining is both an efficient and cost effective approach in saving the cost of peer review and providing a flexible and an impromptu healthcare management mechanism. Further to this, fraud detection is only one part of a bigger program of combating health care fraud, abuse and waste. Fraud detection should note the pitfalls that health care delivery policies like fee for service can create that might increase the possibility of fraud and abuse. The integrated hybrid mining approach to facilitate health insurance payment decision should be covered by diversified decision support shown in **Figure 5**.

Figure 5. Diversified payment decision support

Most of the identified literature focused more on the technical methods used in data mining, and paid little attention to the practical implications of their findings for health care managers and decision makers. Our research provides a good example of a study that provides managerial implications of their findings for dealing with health care fraud. To improve the uptake of hybrid mining methods, future studies should pay more attention to the policy implications.

The mining results can be accumulated and combined with the management strategy. However, it is a passive decision support, because it supports decision processes through new knowledge acquired without producing explicit decision suggestions or solutions. The hybrid mining approach also offers an easy-to-use tool, enabling business users to exploit data mining models with only having a fundamental understanding of the data mining concepts. It also helps them to interpret the models easily. Beginning in September 2014, a Central Intelligence System (CIS) was launched to further screen claimed costs [39]. The system including hybrid mining approach will automatically target important items and picks out anomalies that are then listed for peer review or specially marked, and it also provides information on the anomalies, enhancing the efficiency of the reviews and payment decisions.

Our research also found that most errors or problems of mining happened during the data cleaning and assessment stages, both of which are related to the domain issues. Learning of the health care data analysis process is very time consuming no matter for the students or the researchers, as there are diverse approaches and no one single approach is easy or appropriate at any one time and when an inappropriate method is used, it can lead easily to data misinterpretation. Typically, the health care trial can be segmented into three parts, namely, the structure, process and outcome. Most of the mining projects

ignored the difference among variables relating to the three parts, and simply combine them as a single whole. The result will oftentimes not be useful for the decision support. Besides, the validity of mining would also be put into question, such as in terms of the structure as mistakes are made in choosing variables among the three health care trial stages. The hybrid model is useful to help resolve a number of these noted research and/or application issues. However, it does require more time, additional learning, and/or cost, making it a difficult task to persuade those traditional data analysts who have typically relied on the cause-outcome paradigm to switch their methodological perspective.

In the era of big data, as data sources and technology advance, hybrid mining approach can be used to improve the odds that a certain treatment will result in a favorable outcome for diversified payment decision support. The goal of diversified payment decision support is to use high confidence approach that not only can benefit the cost effectiveness of resources allocation but predict actionable interventions or treatments that will improve long-term health outcomes. Our study explores a hybrid mining approach combined with application system that policy makers should consider using to advance when developing payment policy in health care.

5.2. Limitations and future research recommendations

It is important to note that the content of our research consists of hybrid mining approach in NHI. It does not go into detail results obtained in each payment decision support. This paper includes an exploratory analysis of main characteristic of the payment decision; however, additional aspects could be included and will be considered as future work. The use of National Health Insurance based data is necessary to be careful the bias and confounding issues. Potential Problem of information bias is possible to be happen as the claim is upcoding. The data is also lack of potential confounders to identify the decision as it doesn't cover whole patient medical records. Model need to adjust continuously in order to improve its cost effectiveness and robust methods need to used to account for hybrid mining approach.

Our study explained and demonstrated how to improve the data mining projects through the use of the hybrid mining approach which was composed of design, domain and assessment levels. Accordingly, the design process should synchronize with the other two in order to ensure it can be implemented in the real world. This research also showed that data mining techniques can support the design of future information system for managing and building up the health care knowledge through the analysis of health insurance reimbursement data. The framework presented also illustrates how the hybrid model can help turn data into knowledge. It is

hoped that the results we obtained from using hybrid mining will also encourage other users in applying the framework for future research or cases.

Extending the integration of different methods and the application of hybrid technologies will broaden our understanding on this emerging topic. In future case applications, other quality indicators should be analyzed to effectively support a hospital-wide continuous quality improvement activity and the decision support system should be well integrated with the hospital order communication system to support concurrent review. It is anticipated that new applications of hybrid mining in health care data can be achieved for such areas as the hospital infection control, the ranking of hospitals, or the identifying of high-risk patients. In this sense, the hospital information management system will be developed within a knowledge management context, in which users can share and use the knowledge mode effectively. The hybrid mining results can also be an effective means of understanding behavioral trend of each physician group, and as building evidence in effective credit scoring models. In fact, the credit scoring model still needs to be verified by department, such as internal medicine in future studies.

Hybrid mining based decision support system is already here and its usage is growing rapidly. While applying mining methods to classify consumers' behavior in choosing hospitals, the evidence should be adequate to suggest that the model is useful in identifying existing patterns among hospitals' consumers. What we need is conscious action, the real action that can deliver what has been envisioned for the society as a whole. Last but not least, social consensus on privacy and security is still very critical. The conflict between the sweeping power of technology to access and assemble personal information about privacy and security is an ever-growing concern.

6. References

[1] Han, J., Kamber, H., and Pei J., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2012.
[2] Gupta, G. K., Introduction to Data Mining with Case Studies, PHI Learning Pvt. Ltd., 2014.
[3] Rupink, R., Kukar, M. and Krisper, M., "Integrating Data Mining and Decision Support through Data Mining Based Decision Support System", Journal of Computer Information Systems, Spring 2007, pp. 89-104.
[4] Sliver, M., Sakata, T., Su, H.C., Herman, C., Dolins, S.B. and O'Shea, M.J., "Data Mining Techniques in a Healthcare Data Warehouse", Journal of Healthcare Information Management, Vol. 15, No. 2, Summer 2001, pp. 155-164.
[5] Lee, T., Liua, C., Kuob, Y., Millsc, M.E., Fonga, J., and Hung, C., "Application of data mining to the identification of critical factors in patient falls using a web-based reporting system", International Journal of Medical Informatics, Vol. 80, No. 2, 2011, pp. 141-150.
[6] Chen, Y., Cheng, C., "Identifying the medical practice after total hip arthroplasty using an integrated hybrid

approach", Computers in Biology and Medicine, Vol. 42, No. 8, August 2012, pp. 826-840.

[7] Cheng, T., "Lessons From Taiwan's Universal National Health Insurance: A Conversation with Taiwan's Health Minister Ching-Chuan Yeh", Health Affairs, Vol. 28, No. 4, July/August 2009, pp. 1035-1044.

[8] Cheng, T., "Reflections on the 20th Anniversary of Taiwan's Single-Payer National Health Insurance System", Health Affairs, Vol. 34, No. 3, March 2015, pp. 502-510.

[9] Chen, T.C., and Tsao, H.L., "Using a Hybrid Meta-Evolutionary Rule Mining Approach as a Classification Response Model", Expert System with Applications, Vol. 36, No. 2, March 2009, pp. 1999-2007.

[10] Kuoa, R., Lina, S. and Shihb, C., "Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan", Expert System with Applications, Vol. 33, No. 3, October 2007, pp. 794-808.

[11] Chu, B.H., and Tsai, M.S., "Toward a Hybrid Data Mining Model for Customer Retention", Knowledge-Based Systems, Vol. 20, No. 8, Dec. 2007, pp. 703-718.

[12] Kao, Y.T., Zahara, E., and Kao, I.W., "A Hybridized Approach to Data Clustering", Expert Systems with Applications, Vol. 34, No. 3, April 2008, pp. 1754-1762.

[13] Gopalakrishnan M., Sridhar V., and Krishnamurthy, H., "Some Applications of Clustering in the Design of Neural Networks", Pattern Recognition Letters, Vol. 16, No. 1, 1995, pp. 59-65.

[14] Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., and Lewis, P.A., "Comparative Performance of the FSCL neural net and K-means Algorithm for Market Segmentation", European Journal of Operational Research, Vol. 93, No. 2, Sept. 1996, pp. 346-357.

[15] West, D., "Neural Network Credit Scoring Models", Computers and Operations Research, Vol. 27, No. 11-12, 2000, pp. 1131-1152.

[16] Kuo, R., Ho, L., and Hu, C., "Integration of Self-Organizing Feature Map and K-Means Algorithm for Market Segmentation", Computers & Operations Research, Vol. 29, 2002, pp. 1475-1493.

[17] Lee, T.S., Chiu, C.C., and Chen, I.F., "Credit Scoring Using the Hybrid Neural Discriminant Technique", Expert Systems with Applications, Vol. 23, No. 3, Oct. 2002, pp. 245-254.

[18] Chen, M.C., and Huang, S.H., "Credit Scoring and Rejected Instances Reassigning through Evolutionary Computation Techniques", Expert System with Applications, Vol. 24, No. 4, 2003, pp. 433-441.

[19] Hsieh, N., "Hybrid Mining Approach in the Design of Credit Scoring Model", Expert System with Applications, Vol. 28, No. 4, May 2005, pp.655-665.

[20] Lee, T. and Chen, I., "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines", Expert System with Applications, Available online 11 January 2005.

[21] Huang, C. and Wang, C., "A GA-based feature selection and parameters optimization for support vector machines", Expert System with Applications, Vol. 31, 2006, pp. 231- 240.

[22] Sadatrasoul, M., Gholamian, M., Siami, M. and Hajimohammadi, Z., "Credit scoring in banks and financial institutions via data mining techniques: A literature review", Journal of AI and Data Mining, Vol. 1, No. 2, Spring/Summer 2013, pp. 119-129.

[23] Durairaj, M., and Ranjani, V., "Data Mining Applications in Healthcare Sector: A Study", International Journal of Scientific & Technology Research Vol. 2, No. 10, October 2013, pp. 29-35.

[24] Srinivasan, B. and Pavya, K., "A Study on Data Mining Prediction Techniques in Healthcare Sector", International Research Journal of Engineering and Technology, Vol. 3, No. 3, Mar 2016, pp. 552-556.

[25] Payton, F.C., Data Mining in Health Care Applications, IGI Publishing, 2003.

[26] Hachesu, P., Ahmadi, M., Alizadeh, S. Sadoughi, F., "Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients", Healthcare Informatics Research, Vol. 19, No. 2, June 2013, pp. 121-129.

[27] Syed-Abdul, S., Nguyen, A., Huang, F., Jian, W., Iqbal, U., Yang, V., Hsu, M., and Li, Y., "A smart medication recommendation model for the electronic prescription", Computer Methods and Programs in Biomedicine, Vol. 117, No.2, 2014, pp. 218-224.

[28] ZahidHassan S., and Verma, B., "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases", Proceeding of the Seventh International Conference on Intelligent Systems Design and Applications, 2007, pp. 503-510.

[29] Cabena, P., Hadjinian, P., Stadler, R. and Verhees, J., Discovering Data Mining: From Concept to Implementation, Prentice Hall, 1997.

[30] Chae, Y.M., Ho, S.H., Cho., K.W., Lee, D.H., and Ji, S.H., "Data Mining Approach to Policy Analysis in a Health Insurance Domain", International Journal of Medical Informatics, Vol. 62, 2001, pp. 103-111.

[31] Bohanec, M., Zupan, B., Rajkovic, V., "Applications of Qualitative multiattribute decision Models in Health Care", International Journal of Medical Informatics, 58-59, Sept. 2000, pp. 191-205.

[32] Chen, Y., Yeh, H., Wu. J., Haschler, I., Chen, T., and Wetter, T., "Taiwan's National Health Insurance Research Database: Administrative health care database as study object", Scientometrics, Vol. 86, No. 2, 2010, pp. 365-380.

[33] Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M. and Arab, M.," Using Data Mining to Detect Health Care Fraud and Abuse: A Review of Literature", Global Journal of Health Science; Vol. 7, No. 1; 2015, pp.194-202.

[34] Shina, H., Parkb, H., Lee, J., and Jhee, W.," A scoring model to detect abusive billing patterns in health insurance claims", Expert Systems with Applications, Vol. 39, No. 8, June 2012, pp. 7441-7450.

[35] Lin, K. and Yeh, C.," Use of Data Mining Techniques to Detect Medical Fraud in Health Insurance", International Journal of Engineering and Technology Innovation, Vol. 2, No. 2, 2012, pp. 126-137.

[36] Reinhardt, U.E., "Humbled in Taiwan", BMJ, Vol. 336, Jan. 2008, p.72.

[37] National Health Insurance Administration, National Health Insurance Annual report, 2015-2016.

[38] Yang, W.S., and Hwang, S.Y., "A Process-mining framework for the detection of healthcare fraud and abuse", Expert Systems with Applications, Vol. 31, No. 1, July 2006, pp. 56-68.

[39] Lin, Y., Chen, Y., Hu, S., Chen, H., Chen. J. and Yang, S., "Identifying core herbal treatments for urticarial using Taiwan's nationwide prescription database", Journal of Ethnopharmacology, 2013 July, Vol. 148, No. 2, pp. 556-562.