ICMLA 2011, Honolulu, HI.

Rule-based Prediction of Medical Claims' Payments

A Method and Initial Application to Medicaid Data

Janusz Wojtusiak
George Mason University
Fairfax, VA, USA
jwojtusi@gmu.edu

Che Ngufor
George Mason University
Fairfax, VA, USA
cngufor@gmu.edu

John Shiver
George Mason University
Fairfax, VA, USA
jshiver@gmu.edu

Ronald Ewald
INOVA Health System
Fairfax, VA
ron.ewald@inova.org

***Abstract*—Imperfections in healthcare revenue cycle management systems cause discrepancies between submitted claims and received payments. This paper presents a method for deriving attributional rules that can be used to support the preparation and screening of claims prior to their submission to payers. The method starts with unsupervised analysis of past payments to determine normal levels of payments for services. Then, supervised machine learning is used to derive sets of attributional rules for predicting potential discrepancies in claims. New claims can be then classified using the created models. The method was tested on a subset of Obstetrics claims for payment submitted by one hospital to Medicaid. One year of data was used to create models, which were tested using the following year's data. Results indicate that rule-based models are able to detect abnormal claims prior to their submission.**

Keywords-Rule learning, Hospital billing, Medicaid

I. INTRODUCTION

In healthcare, significant amounts of money are lost due to waste, fraud, and abuse. According to a study released by the American Medical Association, the healthcare system in the United States consumes as much as $210 billion each year on claims processing while as many as one in five claims are processed inaccurately [1]. This includes documentation and revenue cycle management for hospitals, medical group practices, and individual physicians. Across the country, healthcare providers are experiencing ongoing pressure from declining revenues. Payers increasingly need to contain costs. The implementation of healthcare reform through the Patient Protection and Affordable Care Act (PPACA) will exacerbate this issue as it is implemented.

Each payer (government and private) has proprietary clinical documentation standards, service groupings, and client eligibility standards specific to their various product offerings and billing compliance standards. As individual patients have multiple payer coverage and as hospitals and medical providers accept patients with many dozen different payer contracts (each with varying requirements), the management of the billing process and assuring compliance with established standards is complex.

The purpose of this study is to test feasibility of using rule learning to advance healthcare provider revenue cycle management. It uses advanced machine learning algorithms to derive, from historical claims data, support/screening models specific to each payer and insurance product or plan. The models can be subsequently applied to classify new claims. As such, the models are expected to be used to screen every claim for proper documentation prior to submission for payment. In doing so, the provider can prospectively reduce the number and frequency of payment denials for improper claim submissions. Additionally, this methodology can be used to derive more specific models for analyzing reimbursements to match payment for services to each invoice. By manually matching invoices to individual Explanations of Benefits (EOB) the provider can detect payment discrepancies by payers. By utilizing patterns detected from past data and management's experience, information that is indiscernible to any individual using manual techniques, these patterns can be documented and analyzed. This information gives the provider documentable patterns of errors and allow for early corrective actions.

Two ultimate goals of the described method are explored: the ability to predict prior to submission if a specific claim will be processed correctly and receive full payment, or processed incorrectly and declined or not paid in full; and detect regularities in incorrectly processed claims both on the provider and payer sides. In order to achieve these goals, the described method works in three stages: detection of normal payment levels (anomaly detection), creation of rule-based classification, and classification of new claims.

The novelty of the described study, detailed in the following sections, is mainly in focusing on the provider data, understandability and usability of the created models, the ability to deal with both very large and much smaller datasets, and in not using the actual contract information. This sis in contrast to previously performed studies, outlined in the related research section, that focus mainly on the analysis of massive amounts of data, mainly derived from insurance companies, in order to detect fraudulent claims.

The method presented here, can be applied to both large hospitals, as well as smaller providers, including private practices with only few physicians, and clinics. While the presented results focus on Medicaid payments, our current research involves private payers.

II. Data-driven Approach

Billed amounts are determined based on contracts between providers and payers (insurance companies). The contracts specifically define the amounts to be paid for a specific service, or a group of services provided. The calculation of the amounts to be billed for is usually done by software, and sometimes manually in smaller organizations.

The method presented in this section is used to label data for which payments are already received.These are, historical data used for machine learning-based model construction, and newly received payments that are used to update existing models. Creating and updating machine learning models, described later in the paper, requires labeled data with each claim classified as normal or abnormal. The assumption here is that claims that follow payment pattern are normal, and those for which payments do not fit the pattern are abnormal.

The data-driven approach, pursued in the presented study, explicitly ignores contract information. Instead, it discovers amounts that are paid by observing payment trends for a specific payer or a group of payers. For a specific payment received, the method compares its value to previous discovered payments received. If the amounts are the same, the payment is classified as correct. Otherwise, all payments are analyzed within a look-forward window of *k* days, including the day of that claim. If the majority of payments in the window equal the new amount, it is marked as a correct payment, and new level of payment is set with the start date corresponding to the admission date corresponding to the first claim in the window. Otherwise, when the majority of payments are not equal to the analyzed one, it is annotated as abnormal (with additional classification to zero, below normal, and above normal). This method is illustrated in Fig. 1, which shows sample payments received for a subset of Medicaid patients in 2008 for a specific service. In the beginning of July, the amount increased which is detected by the methodology and marked with a horizontal line. This red horizontal line reflects the use of a 30-day look-forward window. Six payments below the July 2008 new normal level of payment are classified as abnormal.

The default look-forward window size is 30 days, however, the window size can be adapted to specific payers and services provided.

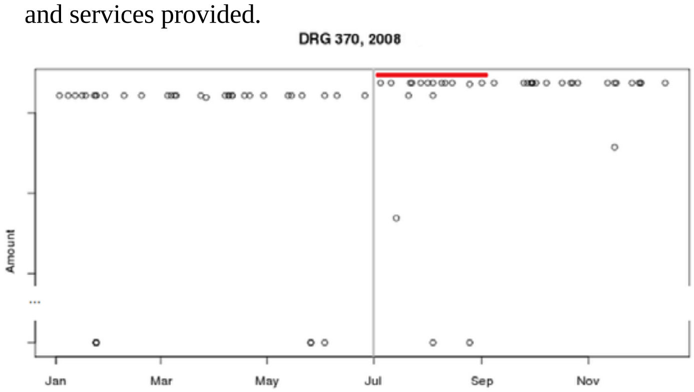

Figure 1. Payments vs. time plot indicating detected contract change date, and look-forward window.

III. System Architecture

The general architecture of the described system consists of six main components depicted in Fig. 2. The process starts with services provided which are coded into ICD-9, DRG, HCPCS and similar nomenclatures. Traditional contract-based claim preparation is performed, which involves software and the knowledge of personnel that prepare bills. At the same time, data-driven method described in the previous section is used to calculate expected payments. Then, claims passed through the rule-based screening module which detects potential discrepancies (Fig. 3). Suspicious claims are marked and passed on for further review before submission as indicated by the backward arrow pointing to contract-based claim preparation. Claims that pass through the rule-based screening are submitted to payers. After payments or denials are received, information is fed back to the rule-based claim classification system which is incrementally updated to account for new situations and the changing environment.

The most important characteristics of the system are:

Combination of data-driven and contract-based claim preparation: There are several reasons for underpayments or claim denials already known by claims management personnel. There is no need to analyze large amounts of data, to find what is already known, but rather focus on rare and unexpected discoveries in data. Thus, access to much smaller amounts of data is needed, and detection of patterns can be done much faster than when searching all patterns.

Rule-based system for classifying claims: The system combines user-defined rules that cover already known cases with machine learning-discovered rules. The rationale for using rules is that they are highly transparent, and can achieve accuracy comparable to the best classifiers (i.e., SVM). Also, rules are used in decision support systems.

The ability to automatically adapt to changing environment: Healthcare environments constantly change, providers' and payers' behaviors change, new contracts are signed on daily bases, and new personnel are responsible for processing claims. The dynamic nature of the problem, calls for methods that can automatically detect changes and adapt to them. Due to availability of incremental machine learning algorithms, the presented architecture allows for doing so.

At the core of the system is the rule-based screening module. Rules in the module are automatically derived from data by a machine learning algorithm, briefly described in the following section. The approach is based on an ensemble of models (classifiers) as depicted in Fig. 3. The models are automatically derived from data, created manually by experts, or by created by combination of the two methods. The rationality behind using combination of automated and expert-based construction of models is to avoid discovering obvious relations in the data (that can be easily described by experts in the form of rules), and on the other hand be able to detect reasons for denials specific for different payers, which may remain unknown even for well-trained individuals or impossible to detect using standard methods.

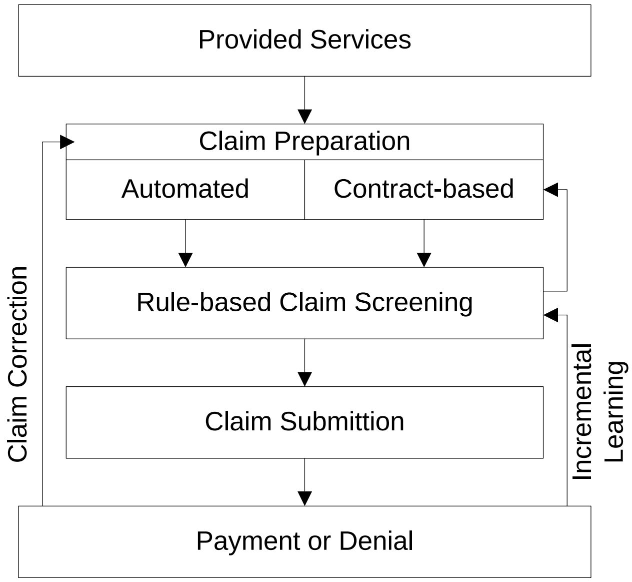

Figure 2. Architecture and data flow.

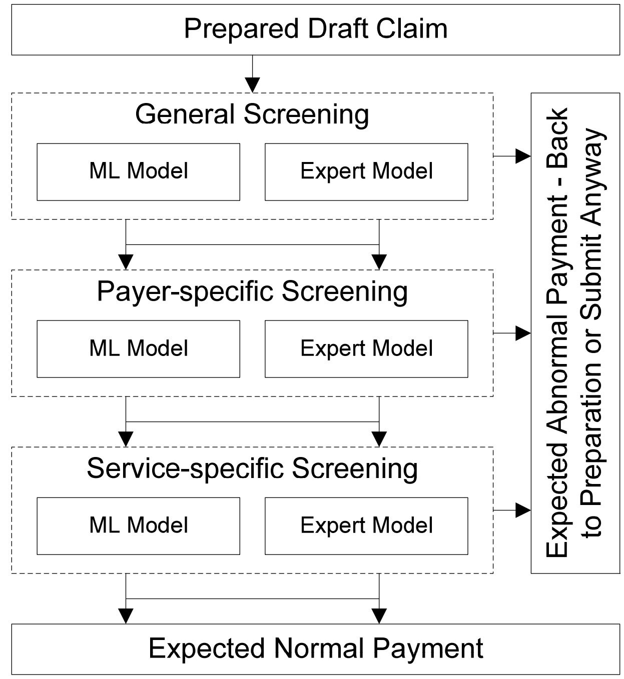

Figure 3. Three levels of bills classification models.

Three levels of models are considered for general screening (for all payers), payer-specific screening, and service-specific screening.

The following steps are used to create models:

Retrieve data: Data are retrieved from a billing system. In the initial study, de-identified billing data for years 2008 and 2009 including both paid and unpaid claims were used. For simplicity, the initial study focused on obstetrical data, and used only Medicaid payments. Work on a discrete database is more likely to produce concrete results in the short period time proposed by this project.

Preprocess data: Data are preprocessed in several steps. First, the data are checked for completeness and consistency. This includes exploratory data analysis needed to gain better understanding of attributes, values, distributions of provided examples, etc. After the check, missing values will be resolved by inserting two basic forms of missing values, namely unknown and not applicable. Next, new attributes are derived from the data to describe high-level billing information and some additional properties such as numbers of empty fields, list empty fields, time from the contract change date, and combinations of values in specific fields. Future work will also include time-based attributes, which are particularly important as they account for previous visits and claims that affect payments. In the presented experiments, data processing has been done using a combination of SQL and R scripts.

Create models: Models can be induced from data, acquired from experts, or created by combining the two.

Learn models from data: State-of-the-art rule-based machine learning software is used to create models from data. The reason for using rule-based models is that they provide a "white box" approach in which learned models can be inspected by human experts and appropriately modified, if needed. The models can also provide useful knowledge to experts. Specifically, this study used the AQ21 machine learning system [13] to derive rules. The system is briefly described in the following section.

Manually acquire models: Simple models can be acquired from human experts to represent basic reasons for denials (i.e. missing key information). Thus, there is no need to analyze massive amounts of data to discover obvious patterns that are well known. To reiterate: the focus of the project is to create models for detecting unexpected patterns of payment that can contribute to experts' knowledge and combined with existing models accurately predict payments.

Apply models: Models are applied in order to categorize previously unseen claims as normal or abnormal. Additionally, abnormal claims can be further categorized as above normal, below normal, and zero. Among claims classified as abnormal, it is possible to perform regression learning in order to determine most likely level of payment. It is, however, important to first perform classification learning and then regression learning.

Test: Models need to be tested before applied in a real world system. Traditionally about 20-30% of data with known outcomes are set aside and used for testing. Due to the relatively small number of abnormal payments and the need to preserve sequence of data, one year of data were used that followed one year of data used for training models.

IV. RULE LEARNING

A. *Why rules?*

Rules are one of the most popular types of knowledge representation used in healthcare. There are also many machine learning algorithms able to derive rules from data.

Rule-based knowledge representation is known to satisfy several criterions in healthcare applications:

Accuracy: Rule-based models can achieve predictive accuracy comparable to other types of models considered in machine learning. Although usually not scoring top in terms of accuracy, rules provide predictions that are good enough for most applications. Accuracy is usually the major criterion considered for quality of learned models.

Transparency: Rule based models are known to be the most transparent and easiest to understand by people not trained in machine learning or statistics. This is particularly important in healthcare where the decision makers (both medical and administrative) need to clearly understand rationale for the decisions being made.

Efficiency: Application of rule-based models is very fast and thousands of rules can be processed every second. This is important in large scale decision support systems operated by multiple users in complex environments.

Transferability: most decision support systems are rule based thus rules resulting from machine learning applications can be used with only minimum change.

B. Attributional Rules

Despite their popularity, standard IF..THEN rules, which use only conjunctions of simple statements, have limited expression power. More expressive form of rules is used in the presented work. Specifically, the main representation of knowledge used in the described method is *attributional rules* [10] whose one form is given by:

CONSEQUENT <= *PREMISE* |_ *EXCEPTION* : *ANNOTATION*

Both *CONSEQUENT* and *PREMISE* are conjunctions of attributional conditions in the form:

[*L REL R* : *A*]

The symbols <=, and |_ denote implication and exception operators, respectively. *EXCEPTION* is either an exception clause in the form of a conjunction of attributional conditions or an explicit list of examples constituting exceptions to the rule. *ANNOTATION* is an additional statistical description, including, for example, the rule's coverage.

An attributional condition corresponds to a simple natural language statement. Its general form is shown above, in which *L* is an attribute, a counting attribute (derived from other attributes), or a simple arithmetical expression over numerical attributes; *R* is an attribute value, internal disjunction or conjunction of values, a range, or an attribute; *REL* is a relation applicable to *L* and *R*; and *A* is an optional annotation that provides statistical information characterizing the condition. The annotation includes numbers of cases satisfied by the condition and its consistency. When *L* is a binary attribute *REL* and *R* may be omitted. Several other forms of attributional rules are available, all of which resemble statements in natural language, and thus are interpretable by people not trained in machine learning [10].

Rule learning usually results in more than one rule outputted by the system. In the investigated approach the focus is on *independent* rules, in which the fact that one rule "fires" does not tell anything about other rules, i.e., the rules do not need to be evaluated in a sequence. In attributional calculus, a set of rules with the same *CONSEQUENT* is called a *ruleset*. Rules in a ruleset represent different reasons for classifying to the same category. For example, there may be three different rules that all classify claims as potentially abnormal. A *ruleset family*, sometimes called a classifier, is a set of rulesets that span over all possible categories in data.

C. Learning Algorithm

In the presented study we used AQ21 software [13] to learn attributional rules for predicting claims' payments. The software is the newest member of the very successful family of AQ learning programs [9][12] that provide users with high flexibility and applicability to a wide range of problems. AQ programs create rules by sequentially covering examples from a given class and avoiding examples from all other classes. This is done by executing a set of logical operators accompanied by statistical rule quality measures, and rule simplicity measures. AQ learning allow for batch rule generation from historical data, as well as incremental modification of existing rules when new data are available.

The AQ21 system is highly configurable and robust system, with features specifically useful in learning form healthcare data. Some of the most important features include the ability to: learn from multitype data (nominal, ordinal, structured (a.k.a. hierarchical), set-valued, interval, ratio, and set-valued attributes); automatically improve representation space through constructive induction; deal with noise in the data; handle unknown, not-applicable, and irrelevant meta-values; learn unordered, structured, or linearly-ordered rulesets; learn from aggregated data and published results; use background knowledge; deal with very small and very large datasets; and generate natural language output. More detailed description of the AQ21 system is available in [13].

D. From Rules to Decision Support

Because rules created by the system are independent (i.e. unordered), they can be easily incorporated into decision support systems. For example, attributional rules described above can be directly written in ARDEN syntax [6]. The actual rules are written in the "logic" slot of Medical Logic Modules (MLMs) and the "data" slot is used to derive attributes' values and translate them into the required format. Because one MLM corresponds to a complete decision, it included multiple rules forming a complete ruleset family. Attributional rules can be also manually inspected by experts and modified as rules and compliance requirements change.

V. Data

The data used in the presented study were derived from the HealthQuest hospital billing system and initially pre-processed using R scripts. The data were then imported to an SQL database for further pre-processing, and then prepared for the AQ21 system that created rule-based models.

Data tables represent patients' demographics, clinical (hospital) information, insurance, and charges. The total number of attributes in the data is 55. The original data consisted of 26,689 records in the demographics and hospital tables, and 30,449 records in the insurance table.

The first task in data preprocessing was to identify Medicaid patients in the data. It is particularly important because some patients are double-eligible and their financial status in the system may not reflect this fact. Thus, in addition to eligibility criteria, it was necessary to select all patients for who Medicaid claims were present. For all identified Medicaid patients, claims were retrieved.

After selecting Medicaid patients, the method described in Section 2 has been used to calculate "normal" values of payment and dates on which the normal values change. For Medicaid payments values of payment correspond to Diagnosis Related Groups (DRGs) – the basis for payment.

For each patient with a Medicaid claim, the total received amount is calculated. The amount is then compared to the normal payment and claims are classified as zero, below normal, normal, and above normal.

For the initial analysis we selected 23 attributes: age, marital status, city, county, state, zip code, employer status code, payor name, admission date, length of hospital stay, admission source, admission type, DRG, ICD-9-CM diagnostic code, ICD-9-CM procedure code, contract identification, covered charges, non-covered charges, covered days, deductible, coinsurance, paid amount, and contractual adjustment. After further elimination a subset of 14 attributes was selected for the rule learning algorithm. This subset of attributes come as a result of performing data-quality checks such as correlation, outlier detection and examining the predictive power of the attributes. Also attributes whose values are not known prior to claims' submission were eliminated.

The subset contained a total of 972 Medicaid records in the 2008 training set and 1005 in the 2009 test set (Table I).

Finally, the data are exported from the SQL database into a single flat text file, which can be uploaded to AQ21 and other machine learning programs.

TABLE I: Distribution of Classes in Training and Testing data.

Data	Zero	Below Normal	Normal	Above Normal
2008	38	34	883	17
2009	23	12	939	31

VI. Selected Results

An initial implementation of the described system has been applied off-line to a set of claims and payments. For simplicity, the initial work focused on obstetrics data and only on Medicaid payments. For Medicaid claims, payments strictly depend on patients' Diagnosis Related Group (DRG). DRG is used to classify patients based on diagnoses and services provided and are the basis for Medicaid reimbursement. For example, all women that deliver through cesarean section without complications have DRG 370.

Example payments for patients with DRG 370 in 2008 were presented earlier in Fig. 1. Despite the simplicity of the data, there are cases present where payments deviate from normal. The dataset is a good testbed for the method.

The data were loaded to the AQ21 rule learning system. First, the payer-specific model based on all used Medicaid OB patients was created. Then, specific models for OB-related DRGs were created. An example rule is shown in Fig. 4. It is one of several rules derived from data.

```
[ payment = below_normal,zero ]
<== [ marital_status = S,U,X] &
    [ zip = ZIP1,ZIP1,ZIP3,ZIP4,ZIP5] &
    [ length_stay >= 1] &
    [ admin_type_id = 3] &
    [ contr_id = XX1,XX2,XX3,XX4,XX5,XX6]
          : p = 9, n = 2, q = 0.733, cx = 55
```

Figure 4. Example rule created by AQ21. Information about specific payers, and patient information has been encrypted.

The rule states that the payment is abnormal (zero or below normal) if patient's marital status is S or U or X, patient's zip code is one of the listed codes, length of stay is at least one day, and so on. At the end of the rule the listed numbers show numbers of abnormal and normal payments satisfying the rule, the rule's quality, and the rule's complexity [10]. Commas separating values within conditions represent internal disjunction, for example in the rule CONSEQUENT payment is below normal or zero.

After creating models using 2008 data, they were tested on 2009 data. They were able to detect about 50% of abnormal payments. The models incorrectly classified only between 5% and 30% of normal payments. First, a provider-specific model for Medicaid payment was built and tested. Then service-specific models for patients with different DRGs were constructed and tested. Summary of the results is presented in Table II.

The results indicate that the method is able to detect abnormal payments in hospital claims data. The 50% detection rate, seems relatively low, but is actually good for this specific dataset. Note that the false positive rate is much lower. This result has significant potential impact on financial management, because detection of even half of incorrectly processed claims may lead to large savings.

TABLE II. NUMBERS AND RATES OF ABNORMAL PAYMENTS AUTOMATICALLY IDENTIFIED IN 2009 DATA.

Model	**Detected Abnormalities**	**False positives**
Medicaid	21/35 (60%)	149/939 (16%)
DRG 371	4/8 (50%)	77/330 (23%)
DRG 372	6/11 (55%)	4/72 (5%)
DRG 373	7/16 (44%)	142/472 (30%)

VII. RELATED RESEARCH

Few published papers focus on detection of detection of claims that most likely will be denied. The work most similar to one presented in this paper is by Kumar et al. [8] who used support vector machines to identify such claims. The authors used a large dataset from an insurance company (3.5 million claims with significantly oversampled incorrect ones). In another closely related work, authors described a rule-based approach called "predictive analytics" to classify hospital payments [3]. The presented method created simple IF … THEN rules. It is not clear what rule induction algorithm has been used, and authors focused on much simpler dataset than one presented here.

On the other hand, significant work has been done in the area of statistical methods for fraud detection [2]. The methods can be supervised, unsupervised outlier detection, or combining the two [4]. A comprehensive review of fraud detection in financial data is presented by Ngai et al. [5] and in healthcare [7]. In healthcare, a number of recent works focused on detection of fraud charges, i.e. submitted to Medicare. The methods focus either on identifying fraudulent charges, of fraudulent providers [11].

Additionally several commercial systems focus on fraud detection. Companies such as SAS and Sybase, provide commercial modules for fraud detection.

VIII. CONCLUSION

Machine learning approach to healthcare claims management provides possibility that go beyond traditional information systems solely based on coding of contracts between payers and providers. The demonstrated method is able to detect abnormalities in patterns and predict potential abnormalities in future claims before their submission.

Experimental results performed on the simple case of Medicaid payments, showed that the method is able to detect irregularities in payments. Although the presented results are scalable to much larger datasets (AQ21 has been successfully applied to problems with millions of examples and problems with thousands of variables), the scalability is a less important tissue. The presented work is done from the provider perspective, thus it is unlikely that much larger datasets will be analyzed. To the contrary, it is more important to detect irregularities and learn models from the smallest possible datasets – a task particularly important for smaller providers, including s practices. Due to combining of logic-based and statistical methods, and the background knowledge, AQ21 is particularly suitable for this task.

In addition to the direct application in prediction of payments, the method has also potential secondary implications. Rule-based models discovered by machine learning are easy to understand by inexperienced people and represent patterns in incorrectly processed claims. Thus, by analyzing these regularities, it may be possible to detect regularities in incorrectly processed claims on both provider and payer side. This may lead to improvement in claim processing and potentially to renegotiation or better specification of payor-provider contracts.

ACKNOWLEDGMENTS

This research has been funded in part by Mason-Inova fund grant. Development of the AQ21 system was partially supported by the National Science Foundation grants.

REFERENCES

[1] "AMA 2010 National health insurance report card", *American Medical Association Annual Meeting*, Chicago, IL., June 14, 2010.

[2] R.J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review" *Statistical Science* vol. 17, iss. 3 August 2002, pp. 235-249.

[3] P. Bradley and J. Kaplan, "Turning Hospital Data into Dollars," *Healthcare Financial Management*, 64, 2, 2010, 64-68.

[4] C.S. Hilas, P.A. Mastorocostas, "An application of supervised and unsupervised learning approaches to telecommunications fraud detection, Knowledge-Based Systems," 21, 7, 2008, 721-726.

[5] E.W.T. Ngai, Yong Hu, Y.H. Wong, Yijun Chen, and Xin Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature" *Decision Support Systems: On quantitative methods for detection of financial fraud*, vol. 50, iss. 3, February 2011, pp. 559-569.

[6] G. Hripcsak, "Writing Arden Syntax medical logic modules" *Computers in Biology and Medicine*, vol. 24, iss. 5, 1994, pp. 331-363.

[7] J. Li, K. Huang, J. Jionghua, S. Jianjun, "A survey on statistical methods for health care fraud detection," Health Care Management Science, 11, 2008, 275-287.

[8] M. Kumar, R. Ghani, and Z. Mei. "Data mining to predict and prevent errors in health insurance claims processing" *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '10). ACM, New York, NY, USA, 2010, 65-74.

[9] R.S. Michalski, "On the quasi-minimal solution of the general covering problem" *Proceedings of the V International Symposium on Information Processing (FCIP 69)(Switching Circuits)*, vol. A3, Yugoslavia, Bled, October 8-11, 1969, pp. 125-128.

[10] R.S. Michalski, "ATTRIBUTIONAL CALCULUS: A Logic and representation language for natural induction" *Reports of the Machine Learning and Inference Laboratory, MLI 04-2*, George Mason University, Fairfax, VA, April, 2004.

[11] R.M. Musal, "Two models to investigate Medicare fraud within unsupervised databases" *Expert Systems with Applications*, vol. 37, iss.12, December 2010, pp. 8628-8633.

[12] J. Wnek, K. Kaufman, E. Bloedorn, and R.S. Michalski, "Inductive learning system AQ15c: the method and user's guide" *Reports of the Machine Learning and Inference Laboratory, MLI 96-6*, George Mason University, Fairfax, VA, August, 1996.

[13] J. Wojtusiak, R.S. Michalski, K. Kaufman , and J. Pietrzykowski, "The AQ21 natural induction program for pattern discovery: initial version and its novel features" *Proc. of the 18th IEEE International Conference on Tools with AI*, Washington D.C., Nov. 13-15, 2006.