# A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile

Pedro A. Ortega
Department of Computer Science
University of Chile
Santiago, Chile
E-mail: peortega@dcc.uchile.cl

Cristián J. Figueroa[1]
Department of Business Intelligence
Sonda S.A.
Santiago, Chile
E-mail: cristian.figueroa@sonda.com

Gonzalo A. Ruz
Manufacturing Engineering Centre
Cardiff University
Cardiff CF24 3AA, UK
E-mail: ruzg2@cf.ac.uk

***Abstract*— This paper describes an effective medical claim fraud/abuse detection system based on data mining used by a Chilean private health insurance company. Fraud and abuse in medical claims have become a major concern within health insurance companies in Chile the last years due to the increasing losses in revenues. Processing medical claims is an exhausting manual task carried out by a few medical experts who have the responsibility of approving, modifying or rejecting the subsidies requested within a limited period from their reception. The proposed detection system uses one committee of multilayer perceptron neural networks (MLP) for each one of the entities involved in the fraud/abuse problem: medical claims, affiliates, medical professionals and employers. Results of the fraud detection system show a detection rate of approximately 75 fraudulent and abusive cases per month, making the detection 6.6 months earlier than without the system. The application of data mining to a real industrial problem through the implementation of an automatic fraud detection system changed the original non-standard medical claims checking process to a standardized process helping to fight against new, unusual and known fraudulent/abusive behaviors.**

## I. Introduction

Health care fraud and abuse losses represent tens of billions of dollars each year in many countries [1] [2] [3] [4] [5]. Medical fraud can occur at various levels [4]. According to [6], fraud and corruption in health care industry can be grouped in illicit activities associated to affiliates, medical professionals, staff and manager, and suppliers. Although fraud may not necessarily lead to direct legal consequences, it can become a critical problem for the business if it is very prevalent and if the prevention procedures are not fail-safe [7]. There is a difference between fraud prevention and detection [4]. Fraud prevention describes measures to avoid fraud to occur in the first place. In contrast, fraud detection involves identifying fraud as quickly as possible once it has been committed.

According to the National Health Care Anti-Fraud Association, health care fraud is an intentional deception or misrepresentation made by a person, or an entity that could result in some *unauthorized* benefit to him or his accomplices. Health care abuse is produced when either the provider practices are inconsistent with sound fiscal, business or medical practices, and result in an *unnecessary* cost or in reimbursement of services that are not medically necessary or that fail to meet professionally recognized standards for health care [8]. In order to assure the healthy operation of a health care insurance system, fraud and abuse detection mechanisms are imperative, but highly specialized domain knowledge is required. Furthermore, well-designed detection policies, able to adapt to new trends acting simultaneously as prevention measures, have to be considered. Data mining which is part of an iterative process called knowledge discovery in databases (KDD) [9] [10] can assist to extract this knowledge automatically. It has allowed better direction and use of health care fraud detection and investigative resources by recognizing and quantifying the underlying attributes of fraudulent claims, fraudulent providers, and fraudulent beneficiaries [11]. Automatic fraud detection helps to reduce the manual parts of a fraud screening/checking process becoming one of the most established industry/government data mining applications [7].

In several countries fraudulent and abusive behavior in health insurance is a major problem. Fraud in medical insurance covers a wide range of activities in terms of cost and sophistication [12]. Health insurance systems are either sponsored by governments or managed by the private sector, to share the health care costs in those countries [13] [14]. Chile has been a pioneer in Latin America initiating programs which decentralized primary health care services and encouraged the development of a private health insurance market in the 1980s [15] [16]. Chilean insured workers and their dependents may channel their mandatory 7% health care payroll contributions to either the publicly managed National Health Fund (FONASA) or to one of the private pre-paid health insurance plans called ISAPREs (*Institutos de Salud Previsional*) [17]. The ISAPREs system, modeled closely on US HMOs (Health Maintenance Organization), was created in 1981 [18] [19] with the objective of giving the consumer more choices through a competitive health insurance system and by expanding private provision of health services. Affiliates of private health insurance companies may pay an additional contribution for a specific health plan. The fee determines the level of coverage and a limit over the health care expenses are defined for each plan. In

[1]Corresponding author. Phone: +56-2-6576247; Fax: +56-2-6575160; Address: Teatinos 500, Santiago-Chile.

1990, the government established a public institution called Superintendency of ISAPREs [20] [21] to regulate the private health insurance market. It mediates between consumers and the ISAPREs, regulates the market in order to guarantee all contracts, and provides information to the public to increase the members' knowledge and market transparency [17].

However, fraud and abuse in medical claims have become an important concern within public and private health insurance companies in Chile the last years due to the increasing losses in revenues. In addition, these losses entail an increase in operation costs diminishing the resources for health benefits. Indeed, a growth of the subsidies by labor incapacity and maternity leaves have been registered in Chile. This increase does not match with an epidemiologist change of the population, but with an inadequate use of the health care system [22]. In 1986, a public institution called COMPIN (*Comisión de Medicina Preventiva e Invalidez*) was created by the government for handling complaints of the affiliates against the ISAPREs [21] [23]. COMPIN has the faculty for modifying the ISAPRE's resolution. Doctor-affiliate complicities are difficult to fight. The labor instability generates potential unemployment which motivates an illegal use of these benefits. Objective parameters for the granting of medical claims do not exist. The exponential increase of the number of medical claims complicates an appropriate control. Processing medical claims is an exhausting manual task that involves analysis, checking, resolution and audit of high volumes of medical claims daily within a limited period of three days from their reception. These control activities are done by few medical experts who have the responsibility of approving, modifying or rejecting the subsidies solicited. Therefore, discovering fraudulent and abusive medical claims is not an easy process for humans experts.

Within the Chilean private health insurance market, Banmedica S.A. ISAPRE is the largest company with a 24.2% market share, delivering health care products, services and information to 609,514 beneficiaries [20] who have labor relationships with approximately 52,500 employers. Moreover, Banmedica S.A. has approximately 18,000 medical contracts with providers and payer organizations. In Banmedica S.A. the process for checking medical claims was totally non-standard. For instance, a particular medical claim could be audited in different ways by two experts. One way is that an expert reviews the medical claim using non-constant criteria through time, making it difficult to establish any specific fraudulent pattern. The other way is based on simple historical rules elaborated by themselves using limited knowledge. Therefore, an effective fraud/abuse detection system based on data mining to detect automatically fraudulent and abusive medical claims was proposed and implemented in Banmedica S.A., discovering non-trivial and novelty knowledge and motivating a pro-active anti-fraud culture within the organization to fight against new, unusual and known fraudulent/abusive behavior

In this paper we describe the before mentioned fraud/abuse detection system that utilizes one committee of multilayer perceptron neural networks (MLP) for each one of the entities involved in the fraud/abuse problem: medical claims, affiliates, medical professionals and employers. This divide-and-conquer strategy allows to feedback information over time, combining affiliates', doctors' and employers' behavior.

The paper is organized as follows. Section II delivers a detailed revision of the literature in relation with the data mining-based health care fraud/abuse detection. Section III presents the features of the proposed automatic fraud detection system. Section IV shows some results about the evaluation of the proposed employer sub-model. Finally section V describes the conclusions of this work with some discussions.

## II. Related Work

Phua et. al. [7] highlights fraud committed in insurance industry as one of the most studied in terms of the number of data mining-based fraud detection publications, existing four sub-groups of insurance fraud detection: home, crop, automobile and medical insurances. In [24] an on-line discounting learning algorithm to indicate whether a case has a high possibility of being a statistical outlier in data mining applications such as fraud detection is used for identifying meaningful rare cases in health insurance pathology data from Australia's Health Insurance Commission (HIC). The performance of a k-Nearest Neighbor (kNN) algorithm with the distance metric being optimized using a genetic algorithm was applied in a real world fraud detection problems faced by the HIC [25]. A multi-layer perceptron (MLP) network was trained to classify the practice profiles of samples of medical general practitioners who had been classified by expert consultants into four classes ranging from normal to abnormal profiles at the HIC [26]. A set of behavioral rules based on heuristics and machine learning are used for performing and scanning a large population of health insurance claims in search of likely fraud [27]. The hot spots methodology that entails the use of clustering and rule induction techniques has been used to identify possible frauds in the Australian Governments public health care system, Medicare [28]. Becker et. al. identify the effects of fraud control expenditures and hospital and patient characteristics on upcoding, treatment intensity and health outcomes in the Medicare and Medicaid programs [29].

Cox [30] applied a fraud detection system based on fuzzy logic for analyzing health care provider claims. This fuzzy system uses rules derived from human experts for detecting anomalous behavior patterns. Tasks performed in support of a data mining project for HCFA (Health Care Financing Administration) such as customer discussions, data extraction and cleaning, transformation of the database, and auditing of the data was described in [11] [31]. A data mining framework that uses the concept of clinical pathways (or integrated care pathways) was utilized for detecting unknown fraud and abusive cases in a real-world data set gathered from the National Health Insurance (NHI) program in Taiwan [13]. Another model that uses attributes mainly derived from various expense fields of claims by experts' consultants was

also designed to detect suspicious claims in the Taiwan NHI program [32]. An overview of the merging neural networks, fuzzy logic and genetic algorithms applied to the insurance industry was done in [33]. Neural networks to classify fraudulent and non-fraudulent claims for automobile bodily injury in healthcare insurance claims was implemented in [34]. A method based on naive Bayes that effectively combines the advantages of boosting and the explanatory power of the weight of evidence scoring framework was presented in [35]. The method consists of closed personal injury protection (PIP) automobile insurance claims from accidents that occurred in Massachusetts during 1993 and were previously investigated for suspicion of fraud by domain experts.

A temporal pattern mining algorithms to identify a set of frequent temporal patterns gathering insurance claim instances about Pelvic Inflammatory Disease from regional hospitals in Taiwan was presented [36] [37]. Furthermore, the classification algorithm C4.5 was applied for fraud/abuse detection by using the discovered temporal patterns as predictive features. Some data mining-based approaches which can be used to extract medical knowledge for diagnosis, screening, prognosis, monitoring, therapy support or overall patient management were presented in [38]. Fraud indicators and rules from the knowledge and experience of human experts to develop a computer-based expert to facilitate the work of insurance carriers was carried out by [39]. In Chile, a single neural network to detect fraudulent medical claims was implemented in another healthcare insurance company [40]. This scheme utilizes all the data available in arriving medical claim for constructing a unique vector which is evaluated by the single neural network.

Several actions that employers can take to reduce losses from health care fraud, suggesting policy statements and administrative procedures and guidelines to discourage employee fraud, to combat provider fraud, to improve health care fraud detection through the claims payment system and the possibility of civil and criminal remedies and reviews the legal theories were pointed out in [12]. Trends in published neural network application research and exploration of potential neural network research areas were mentioned in [41]. A short discussion of risk and fraud in the insurance industry was given in [42].

## III. Methods

### A. Entities and Medical Claim Data

A medical claim involves the participation of an affiliate, who is finally the direct benefitted; a medical professional, who has the faculty for fulfilling the requirement of the resting period and subsidies for the affiliate; and an employer, who represents legally the company where the affiliate works.

A private health insurance ISAPRE company like Banmedica S.A. processes daily approximately 800 medical claims. Each claim is submitted by an affiliate under the approval of a medical professional justifying the work incapacity. Data such as age, sex, type of claim, affiliate's name and date of birth, ID number, resting period solicited, type and place of the resting, identification of the medical professional, identification of the employer, labor activity of the company where the affiliate works, affiliate's profession and income records that the affiliate has gotten in the last three months are incorporated in each form.

On the other hand, data such as main and secondary diagnoses, supporting clinical records and medical exams are also fulfilled by the medical professional, but this data is sealed when the form arrives to some of the ISAPRE's agencies. Once ISAPRE's medical experts check the forms they can unveil this data. Just in that moment the medical experts decide on whether to approve, modify or reject the number of resting days solicited. In this sense, a neural classifier that makes a predictive detection of the fraudulent and abusive claims would be of great help for the medical experts in their reviewing process, acting as a pre-screen filter. This predictive detection must only consider historic data associated to the affiliate, the medical professional and the employer, and data available before the medical revision of the arriving medical claim.

### B. Business and Data Understanding

Several meetings were held with some medical experts who explained the main aspects of their criteria for approval, modification and rejection of medical claims. This allowed us to better understand the underlying business model, including discriminative behavioral patterns, as well as weaknesses of the current non-standarized fraud detection procedure. The outcome of these meetings was a preliminary set of variables designed to discriminate between normal and suspicious/fraudulent behavior.

Our analysis started with two sets of medical claims. The first set contained 169 historic abusive medical claims that were richly documented by Banmedica S.A. until 2003. These cases manifested notorious fraudulent patterns along extended periods, hence their manual detection was a simple task for the medical experts. The second set was a sample of 500,000 labelled medical claims recorded between 2001 and 2003. These were either "approved", "rejected" or "reduced".

In order to build an appropriate set for the application of machine learning techniques, both sets had to be completed by computing their remaining descriptive features from the ISAPRE's relational database. In particular, a subset of nine tables was used. These tables had data associated to each entity involved in a medical claim as for instance payment behavior, medical resolutions, frequencies, beneficiaries and historic incomes. Consequently, a single data set was calculated and stored into a unique repository for further data analysis.

Next, several tasks to assess the data quality were done. We checked for correlations among features and discriminative power, as well as their consistency with empirical knowledge. For instance, seasonality of medical claim submission with respiratory, psychiatric, maternal leaves diagnoses, correlations between days submitted and effective days of subsidy and income and payroll contribution analysis were carried out. Finally, a total amount of 125 features were selected.

However, this analysis also showed that the majority of the 169 fraudulent/abusive documented medical claims were concentrated around a reduced number of medical professionals, affiliates and employers. For example, 19 employers and 6 doctors were implicated with 152 medical claims. Furthermore, due to the lack of standards and the huge amount of medical claims, the labelling of the second data set was not sufficiently accurate, and in some cases even contradictory. Moreover, 35% of these medical claims belonging to year 2001 were discarded due to their poor quality in terms of missing values and low contribution. Also, medical claims whose number of solicited resting days was lower than 10 were excluded, according to their almost negligible impact in profits.

The poor quality of the data forced us to rebuild the data set by relabelling a subset of cases and to apply aggressive dimensionality reduction techniques.

### C. Data Preparation

In order to build a robust classifier using only a small training set, it was decided to apply a divide-and-conquer strategy. The initial problem was subdivided into smaller problems, namely, four separate models to cope with the entities before mentioned. Each sub-model required a smaller number of features and training samples. The detailed description follows.

First, an exhaustive manual classification was done, assisted by both medical experts and legal advisors. For each entity, that is, medical claims, affiliates, medical professionals and employers, training sets of sizes 2838, 424, 590 and 394 samples were obtained. Each set had an equal proportion of fraudulent/abusive and normal cases.

Second, the feature vectors for the four different models were further optimized. A manual feature selection procedure was applied to reduce the feature vectors. For example, for groups of highly correlated features, only the most discriminative one was chosen. Furthermore, a categorical feature was avoided or replaced by a continuous feature whenever possible. Intimately related but complementary variables were fused into non-human interpretable features that are better suited for a continuous classifier. The overall maxim has been to design robust features that summarize temporal behavior over an extended time span. Most of them involved historical data, generally moving windows of 12 months backwards starting from the submission date. As a result, the medical claim, affiliate, medical professional and employer sub-models had feature vectors of sizes 14, 25, 17 and 12 respectively. As an example, Fig. 1 shows the histogram of the variable *expected resting days approved by the employer* for the 394 manually classified employer cases, separated by fraudulent/abusive ($F$) and normal ($N$) cases.

Finally, standard data preparation techniques have been applied to avoid training biases. These include the removal of 2% of the outliers and a linear normalization of the features. The data transformation parameters have been saved for their later use during production stage.

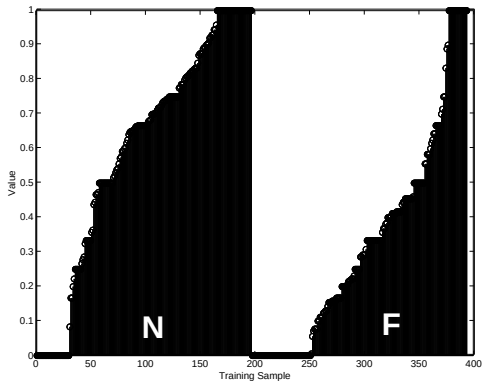


Fig. 1. Histogram of the variable *expected resting days approved by the employer* for the 394 manually classified employer cases, separated by fraudulent/abusive ($F$) and normal ($N$) cases

### D. Modelling

The proposed system allows the identification of fraudulent and abusive behavior in medical claims, affiliates, medical professionals and employers. Upon the medical claim submission, the different entities are analyzed separately using historical data with cross-references among them. As mentioned above, the predictive model cannot make use of the sealed diagnostic information contained in the medical claim for its decision, although historical data is incorporated via the features.

TABLE I
EMPLOYER MODEL, CLASSIFICATION VARIANCE

| Classifier | Mean | Std. Dev. |
|---|---|---|
| Single MLP | 88.7 % | 8.4 % |
| Committee (10 MLPs) | 88.7 % | 1.8 % |

Initially, each sub-problem was modelled by a standard two-layer neural network (MLP). Albeit architectures were kept simple (e.g. using small hidden layers), the classifiers' accuracy showed a surprisingly high variance that exceeded 8% between different runs. In order to decrease the model variance, we decided to implement a *committee* of multilayer feedforward neural networks [43][44]. Alternative complexity reduction techniques such as network pruning were discarded due to the high weight variance even for very simple architectures. In the case of committees, the output of several independently trained networks was averaged in order to reduce their model variance, thus delivering robust results specially when the number of samples is low. Each sub-model had committees of 10 MLPs. Table I compares the standard deviation obtained for single network against a committee. The chosen architectures were 14-3-1, 25-3-1, 17-3-1 and 12-3-1 for the medical claim, affiliate, medical professional and employer committees respectively. All activation functions were sigmoidal. The output unit indicates whether the sample corresponds fraudulent/abusive or normal behavior. For each case, the labelled data set were divided

into a training set, a validation set and a testing set. The training procedure minimized the error measured on the validation set while using the training set to adjust the networks' weights. This technique known as *early stopping* provides a principled method for selecting models that generalize well without sacrificing capacity, hence avoiding over-fitting to the training data's noise while keeping the classifier's ability to learn non-linear discriminant boundaries [45]. The testing sets were used to estimate the generalization performance of the system. It is worth noticing that the design strives for model simplicity and improved generalization ability under a limited sample set scenario.

Our fraud detection system is illustrated in Fig. 2. This diagram shows the four sub-models, namely, a committee of neural networks for each entity. In each case, the inputs are precomputed feature vectors that model the particular fraud and abuse subproblem. The output of each committee delivers a predictive value each time that a medical claim is received by the ISAPRE. These values serve as additional inputs to the sub-models, providing a feedback mechanism for combining the different results. For instance, the outputs of the affiliate, medical professionals and employer models provide strong and condensed evidence for the medical claim sub-model. Models are evaluated at fixed time intervals according to a predefined schedule. The medical claim model is executed daily to process incoming submissions, whereas the other are executed once monthly.

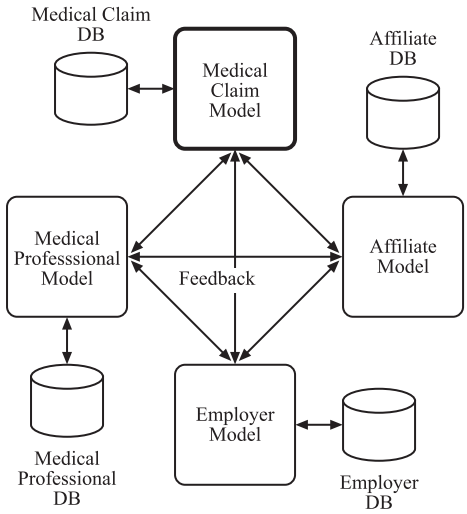


Fig. 2. Fraud detection scheme by using sub-models with feedback connections.

All sub-models are retrained monthly. In order to keep the training samples representative for historical and novel fraud behavior, a data renewal procedure has been defined. New training samples are chosen randomly and rigorously classified by experts. A subset of equal proportions of fraud and normal cases are then selected and incorporated to the training set. This way, the model is kept updated and informed of novel fraud trends, acting simultaneously as a prevention mechanism.

### *E. Incorporation into Fraud Detection Workflow*

Approximately 800 digitalized medical claim forms arrive during each day. Our automated fraud detection system is executed at night, assigning a *fraud probabilities* to each form. Associated affiliates, medical professionals and employers are updated as well. At the next day, a web interface allows to consult these records ranked by their fraud probabilities, acting like a pre-screen filter.

Several costs are involved in this fraud detection procedure, such as the personnel's salaries, false alarms and fraud or abuse cases that are not detected. Based on this information, an optimum fraud probability threshold can be estimated. Records with lower fraud probabilities can be excluded from this revision procedure. The description follows.

As mentioned in [46], a naive classifier can get high accuracy on a skewed data set, where the prevalence of fraud is very low, but at the cost of misclassifying all the true fraudulent/abusive cases. Therefore, an optimum decision threshold was estimated by plotting receiver operating characteristic (ROC) curves [47] instead of the classifier accuracy. ROC curves plot the true positive (TP) against the false positive (FP) rates at different decision making thresholds and determine the test's ability to differentiate between fraudulent/abusive and normal cases. TP and FP are the per unit rate of correct and false fraudulent/abusive classifications respectively. Moreover, an iso-performance line was plotted in the $TP-FP$ plane. Classifiers on this line all have the same expected cost [48]. Two points $(TP_1, FP_1)$ and $(TP_2, FP_2)$ have the same performance if

$$\frac{TP_2 - TP_1}{FP_2 - FP_1} = \frac{p(N)C(FP)}{p(P)C(FN)}, \tag{1}$$

where $p(N)$ and $p(P)$ are the prior probabilities of obtaining a negative and a positive example respectively, and $FN$ is the false negative rate. The $C(FP)$ and $C(FN)$ represent the costs of a $FP$ and a $FN$ error, respectively. This equation defines the slope of the iso-performance line. Given a ROC curve and an iso-performance line, the point of intersection defines the optimal operation point.

## IV. Results

The results associated with our fraud detection system such as classifier accuracy and total savings are confidential due to a disclosure contract. Nevertheless, Banmedica S.A. has granted the permission to publish results related to the employer sub-model.

Historical data considering a total of 8,819 employers was analyzed. This set contains 418 fraudulent/abusive and 8,401 normal cases. This data set was divided into a training, validation and test set (Table II). For training purposes, the fraud/abuse cases in both the validation and training sets were quadruplicated to remove possible bias.

Figure 3 shows the ROC curve obtained with the prediction employer sub-model. The top left point (0,1) represents

TABLE II
EMPLOYER DATA SET

| Employer Category | Number of Cases | Percentage |
|---|---|---|
| Fraud/abuse (T) | 176 | 20.0 |
| Normal (T) | 706 | 80.0 |
| Total (T) | 882 | 100.0 |
| Fraud/abuse (V) | 118 | 20.1 |
| Normal (V) | 470 | 79.9 |
| Total (V) | 588 | 100.0 |
| Fraud/abuse (T) | 124 | 1.7 |
| Normal (T) | 7,225 | 98.3 |
| Total (T) | 7,349 | 100.0 |

the perfect predictor where all fraudulent/ abusive cases are caught without generating false alarms. Therefore, the closer the ROC curve is to the point (0,1) the better is the performance. To plot the iso-performance line, the prior probabilities have been estimated using random sampling and manual inspection (based on data between June 2005 and January 2006), obtaining $p(P) = 0.017$ and $p(N) = 0.983$. Furthermore, a scenario where the cost of a $FN$ error corresponds to at least 5 times the cost of a $FP$ error has been setup, yielding a cost ratio of $C(FP)/C(FN) > 0.2$. At the operation point the TP and FP rates were 73.4% and 6.9%, respectively, i.e., the prediction module was able to identify 73.4% of the true fraudsters/abuses, screening only 8.7% of all employers. Although there were 6.9% false positives, this is acceptable for a model which is designed to work as a pre-screen filter. Those identified as fraudsters/abuses by the committee are then referred for a detailed revision.

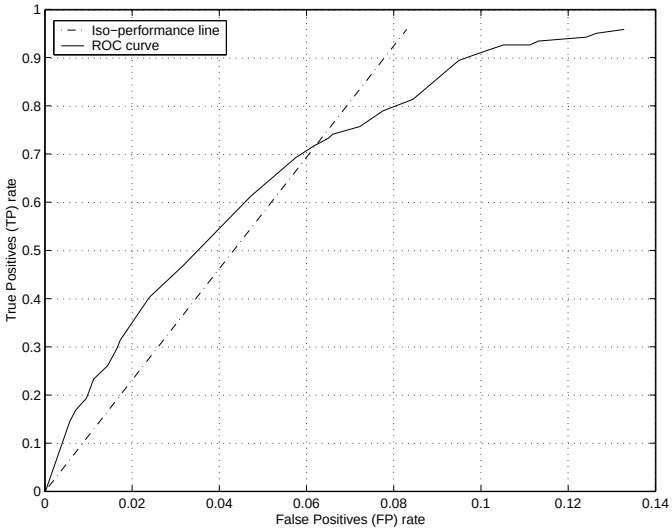


Fig. 3. ROC curve

To illustrate the efficacy of our fraud detection system, Table III shows the evolution of an employer's fraud score. In this case, an employer whose first record in the ISAPRE is saved on September 13, 2005 is analyzed. The employer ID number is 7771xxxx. This first record originated from an associated affiliate's medical claim submission. The predictive employer sub-model was only able to deliver a value 0.61 of being fraudulent or abusive because there is no history associated to the employer until that point. The second medical claim arrived to the ISAPRE on October 13, 2005. After the update, employer sub-model scored this record with an abuse probability of 0.69. Although this score increased since the first submission, it did not reach the decision threshold. After a third submission on October 18 and a fourth on October 31, the score augmented to 0.82 and 0.91 respectively, triggering its manual revision procedure. The experts confirmed the abuse case and activated an alarm rule that highlights future medical claims from this employer. Consequently, our fraud detection system was able to identify this case within only 2 months. This single example is representative: based on historical data, our fraud detection scheme identifies fraud/abuse employer behavior within 2 months, whereas the former procedure took 8.6 months in the average.

TABLE III
FRAUD SCORE EVOLUTION, EMPLOYER ID 7771XXXX

| Fraud/abuse probability | Reception date of medical claim |
|---|---|
| 0.6118 | September 13, 2005 |
| 0.6908 | October 13, 2005 |
| 0.8246 | October 18, 2005 |
| 0.9126 | October 31, 2005 |

## V. CONCLUSIONS AND DISCUSSION

With the implementation of our fraud detection system in medical claims, a proactive anti-fraud culture was generated within Banmedica S.A. to fight against new, unusual and known bad behaviors filtering opportunely suspicious medical claims. Insightful knowledge has been gathered, allowing to build a fraud/abuse taxonomy that identifies 15 different types. Moreover, this automatized revision procedure motivated improvements in the manual revision process.

During 2005 Banmedica S.A. was able to estimate that the new fraud detection system rejected medical claims that contributed between 9.5% and 10.0% of the overall raw costs. Additionally, persistent fraud and abuse cases are detected 6.6 months earlier than without the system, as estimated from historical records. The savings are considerably higher due to the premature detection of these cases

Banmedica S.A. has changed its fraud detection policy from its former reactive detection strategy, which yielded only a few documented cases up to 2003, to a proactive and preventive detection strategy, reaching approximately 75 fraudulent and abusive cases each month during 2004 and 2005. This drastic increase is mainly due to the contribution of our fraud detection system. The current fraud detection rate yields savings that cover operational costs and allowed

to increase the quality of the health care coverage, fully justifying the investment.

Unlike other fraud detection systems that are based on a single monolithic model, our divide-and-conquer approach is able to analyze each involved entity separately, along with other beneficial side-effects such as dimensional reduction and the consequent model robustness. It is worth noticing that the detection of the sources of fraudulent and abusive behavior is a far more efficient strategy than the analysis of the individual medical claims. The detection of a source, namely, an affiliate, a medical professional or an employer, allows to cover multiple present and future submissions.

Future work considers the design of an improved model to estimate costs and savings associated to the operation of our detection scheme, as well as to measure its preventive secondary effects.

## ACKNOWLEDGMENT

The authors would like to thank Arturo Phillips, Jaime Ochagavía and Verónica Rodríguez who belong to the health management of Banmedica S.A. for their valuable help and support during the implementation of the model. We also thank Johanna Guzmán, Víctor Marchant, Cristián Ríos, Juan Pablo Herrera, Mauricio Marcos and Andrés Vergara who belong to the Business Intelligence Department of Sonda S.A. for their help in building the repository and managing different activities that allowed the proper installation of the fraud detection system.

## REFERENCES

[1] United States General Accounting Office, "Health insurance: Vulnerable payers lose billions to fraud and abuse," GAO-HRD-92-69, 1992.

[2] M. Lassey, W. Lassey and M. Jinks, "Health care systems around the world: Characteristics, issues, reforms," *Englewood Cliffs*, NJ: Prentice-Hall, 1997.

[3] S. J. Hong and S. M. Weiss, "Advances in predictive models for data mining," *Pattern Recognition Letters*, vol. 22, pp. 55–61, 2001.

[4] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, no. 3, pp. 235–249, 2002.

[5] Health care fraud, "Health care fraud: A serious and costly reality for all Americans," National Health Care Anti-Fraud Association (NHCAA), REPORT, http://www.nhcaa.org, April 2005.

[6] European healthcare fraud and corruption network, http://www.ehfcn.org/, 2005.

[7] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research", Artificial Intelligence Review, submitted, 2005.

[8] Guidelines to health care fraud, "Guidelines to health care fraud," *National Health Care Anti-Fraud Association (NHCAA)*, NHCCA Board of Governors, http://www.nhcaa.org, 1991.

[9] U. M. Fayyad, "Data mining and knowledge discovery: making sense out of data," IEEE Expert, Intelligent Systems and their Applications, pp. 2025, October 1996.

[10] J. Han and M. Kamber, Data mining: Concepts and techniques, Morgan Kaufmann Publishers, San Francisco, 2001.

[11] L. Sokol, B. Garcia, J. Rodriguez, M. West and K. Johnson, "Using data mining to find fraud in HCFA health care claims," *Top Health Information Management*, vol. 22, no. 1, pp. 1-13, 2001.

[12] B. B. Pflaum and J. S. Rivers, "Employer strategies to combat health care plan fraud," *Benefits Quarterly*, vol. 7, no. 1, pp. 6-14, 1991.

[13] W. S. Yang and S. Y. Hwang, "A process-mining framework for the detection of healtcare fraud and abuse," *Expert Systems with Applications*, Article in Press, corrected proof, 2005.

[14] L. J. Opit, "The cost of health care and health insurance in Australia: Some problems associated with the fee-for-service," Soc. Sci. Med. vol. 18, no. 11, 967–972, 1984.

[15] J. Jimenez and T. Bossert, "Chile's health sector reform: lessons from four reform periods," *Health Policy*, vol. 32, pp. 155–166, 1995.

[16] A. Viveros-Long, "Changes in health financing: The Chilean experience," Soc. Sci. Med. vol. 22, no. 3, pp. 379–385, 1986.

[17] F. Bertranou, "Are market-oriented health insurance reforms possible in Latin America?. The cases of Argentina, Chile and Colombia," *Health Policy*, vol. 47, pp. 19–36, 1999.

[18] E. Miranda, J. Scarpaci and I. Irarrázaval, "A decade of HMOs in Chile: market behavior, consumer choice and the state," *Health & Place*, vol. 1, pp. 51–59, 1995.

[19] C. Sapelli and B. Vial, "Self-selection and moral hazard in Chilean health insurance," *Journal of Health Economics*, vol. 22, pp. 459–476, 2003.

[20] Superintendencia de ISAPRES, October, 2005 http://www.superintendenciadesalud.cl/.

[21] Ley N 18.933 Publicada en el Diario Oficial de 09.03.90 Crea Superintendencia de ISAPRE, dicta normas para el otorgamiento de prestaciones por ISAPRE y deroga el D.F.L. N 3, de 1981, del Ministerio de Salud, Chile.

[22] P. Melero, "Cara Evidente de Licencias Médicas: Incentivos Invertidos," Comisión de Salud, Cámara de Diputados de Chile, Agosto, 2005, available at http://www.isapre.cl.

[23] Decreto supremo N 42, 1986, Ministerio de Salud, Chile.

[24] K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," In Data Mining and Knowledge Discovery vol. 8, pp. 275–300, 2004.

[25] H. He, W. Graco, and X. Yao, "Application of genetic algorithms and k-nearest neighbour method in medical fraud detection," In Proc. of SEAL1998, pp. 74–81, 1999.

[26] H. He, J. Wang, W. Graco and S. Hawkins, "Application of neural networks to detection of medical fraud," *Expert Systems with Applications*, vol. 13, no. 4, pp. 329–336, 1997.

[27] J. Major, and D. Riedinger, "EFD: A hybrid knowledge/statistical-based system for the detection of fraud," Journal of Risk and Insurance 69(3), pp. 309-324, 2002.

[28] G. Williams, "Evolutionary hot spots data mining: An architecture for exploring for interesting discoveries," In Proc. of PAKDD99, 1999.

[29] D. Becker, D. Kessler and M. McClellan, "Detecting medicare abuse," *Journal of Health Economics*, vol. 24, pp. 189–210, 2002.

[30] E. Cox, "A fuzzy system for detecting anomalous behaviors in healthcare provider claims," In Goonatilake, S. & Treleaven, P. (eds.) Intelligent Systems for Finance and Business, pp. 111-134. John Wiley and Sons Ltd., 1995.

[31] L. Sokol, B. Garcia, M. West, J. Rodriguez and K. Johnson, "Precursory steps to mining HCFA health care claims," *Proceedings of the Hawaii International Conference on System Sciences*, Hawaii, pp. 6019, 2001.

[32] C. L. Chan and C. H. Lan, "A data mining technique combining fuzzy sets theory and Bayesian classifier – An application of auditing the health insurance fee," *Proceedings of the International Conference on Artificial Intelligence*, pp. 402–408, 2001.

[33] A. F. Shapiro, "The merging of neural networks, fuzzy logic, and genetic algorithms," *Insurance: Mathematics and Economics*, vol. 31, pp. 115–131, 2002.

[34] P. L. Brockett, X. Xia, and R. A. Derrig, "Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud," The Journal of Risk and Insurance, vol. 65, no. 2, pp. 245–274, 1998.

[35] S. Viaene, A. Richard and D. G. Dedene, "A case study of applying boosting Naive Bayes to claim fraud diagnosis," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 5, pp. 612–620, May 2004.

[36] C. P. Wei, S. Y. Hwang and W. S. Yang, "Mining frequent temporal patterns in process databases," Proceedings of International Workshop on Information Technologies and Systems, Australia: Brisbane, pp. 175-180, 2000.

[37] S. Y. Hwang, C. P. Wei, and W. S. Yang, "Process mining: Discovery of temporal patterns from process instances," *Computers in Industry*, vol. 53 no. 3 pp. 345-364, 2004.

[38] N. Lavrac and N. Lavrac, "Selected techniques for data mining in medicine," *Artificial Intelligence in Medicine*, vol. 16 no. 1, pp. 3-23, 1999.

[39] W. Herb and M. Tom, "A scientific approach for detecting fraud," Best's Review, vol. 95, no. 4, pp. 78–81, 1995.

[40] C. Cooper, "Turning information into action," Computer Associates: The Software That Manages eBusiness, Report, January 2003, available at http://www.ca.com.

[41] B. K. Wong, T. A. Bodnovich and Y. Selvi, "Neural network applications in business: A review and analysis of the literature (1988-95)," *Decision Support Systems*, vol. 19, pp. 301–320, 1997.

[42] B. Glasgow, "Risk and fraud in the insurance industry," AAAI Workshop: AI Approaches to Fraud Detection and Risk Management, AAAI Press, Menlo Park, California, pp. 20–21, 1997.

[43] M. P. Perrone, "General averaging results for convex optimization", *Proceedings 1993 Connectionist Models Summer School*, pp. 364-371, Hillsdale, NJ: Lawrence Erlbaum, 1994.

[44] M. P. Perrone, L. N. Cooper, "When networks disagree: ensemble methods for hybrid neural networks", *Artificial Neural Networks for Speech and Vision*, pp. 126-142, London: Chapman & Hall, 1993.

[45] R. Caruana, S. Lawrence, C. L. Giles, "Overfitting in neural networks: backpropagation, conjugate gradient, and early stopping", *Neural Information Processing Systems*, Denver, Colorado, 2000.

[46] P. A. Estévez, C. M. Held and C. A. Perez, "Subscription fraud prevention in telecommunications using fuzzy rules and neural networks," *Expert Systems with Applications*, vol. 31, no. 2, 2006, (in press).

[47] T. Fawcett, "ROC graphs: Notes and practical considerations for data mining researchers," HP Labs. Technical Report HPL-2003-4, 2003.

[48] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance with nonuniform class and cost distributions," *AI Approaches to Fraud Detection & Risk Management*, Workshop Technical Report WS-97-07, AAAI Press, Menlo Park, California, 1997.