Journal of Network and Innovative Computing
ISSN 2160-2174, Volume 1 (2013) pp. 341-347

Computational Intelligence Models for Insurance Fraud Detection: A Review of a Decade of Research

Amira Kamil Ibrahim Hassan[1] and Ajith Abraham[2]

[1]Department of Computer Science, Sudan University of Science and Technology, Sudan Khartoum, Sudan
amirakamil2@yahoo.com
[2] Machine Intelligence Research Labs (MIR Labs), WA, USA
ajith.abraham@ieee.org

Abstract: **This paper presents a review of the literature on the application of data mining techniques for the detection of insurance fraud. Academic literature were analyzed and classified into three types of insurance fraud (automobile insurance, crop insurance and healthcare insurance) and six classes of data mining techniques (classification, regression, clustering, prediction, outlier detection, and visualization). The findings of this review clearly show that automobile insurance fraud detection have also attracted a great deal of attention in recent years. The main data mining techniques used for insurance fraud detection are logistic models, Decision tree, the Naïve Bayes, and support vector machine.**

Keywords: insurance fraud, fraud detection, data mining**.**

I. Introduction

Insurance fraud is a significant and costly problem for both policyholders and insurance companies in all sectors of the insurance industry. In recent years, fraud detection has attracted a great deal of concern and attention. The Oxford English Dictionary [1] defines fraud as “wrongful or criminal deception intended to result in financial or personal gain”.

Fraud occurs in a wide variety of forms and is ever changing as new technologies and new economic and social systems provide new opportunities for fraudulent activity. The total extent of business losses due to fraudulent activities is difficult to define.

Phua et al. [2] described fraud as leading to the abuse of a profit organization's system without necessarily leading to direct legal consequences. Although there is no universally accepted definition of financial fraud, Wang et al. [3], defined it as “a deliberate act that is contrary to law, rule, or policy with intent to obtain unauthorized financial benefit”. Economically, insurance fraud is becoming an increasingly serious problem.

According to a 2007 BBC news report [4], fraudulent insurance claims cost UK insurers a total of 1.6 billion pounds a year. The overall losses caused by insurance fraud are incalculable. Insurance fraud detection is important for preventing the disturbing results of insurance fraud. IFD involves distinguishing fraudulent claims from genuine claims, thereby disclosing fraudulent behavior or activities and enabling decision makers to develop appropriate strategies to decrease the impact of fraud.

Data mining has a significant role in IDF, as it is often applied to extract and uncover the hidden truths behind very large quantities of data. Data mining is about finding insights which are statistically reliable, unknown previously, and actionable from data [5]. This data must be available, relevant, adequate, and clean. Also, the data mining problem must be well-defined, cannot be solved by query and reporting tools, and guided by a data mining process model [6].

Bose and Mahapatra [7] defined data mining as a process of identifying interesting patterns in databases that can then be used in decision making. Turban et al. [8] defined data mining as a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and subsequently gain knowledge from a large database. Frawley et al. [9] stated that the objective of data mining is to obtain useful, non-explicit information from data stored in large repositories. Kou et al. [10] highlighted that an important advantage of data mining is that it can be used to develop a new class of models to identify new attacks before they can be detected by human experts. Phua et al. [11] pointed out that fraud detection has become one of the best established applications of data mining in both industry and government. Various data mining techniques have been applied in IFD, such as neural networks, logistic regression models, the naïve Bayes method, and decision trees, among others.

II. Methodology of the Research

The methodology used in this paper consists of three stages [12]. In the first stage the research area, aim and scope are defined. The data mining techniques used for insurance fraud detection in published academic papers is the research area of this paper. The aim of this paper is to classify the papers according to the used data mining techniques. The scope is all published papers on insurance fraud detection (IFD) using data mining technique in the period between 1997 and 2013. This topic of research is relatively new so the 16 year period of this study is considered a good indicator for the data mining techniques used to IFD.

In the second stage the search and selection criteria is specified. Also classification framework is build for the selected papers. The search was done in five different academic databases. A complete list of published papers in this area was formed.

The search phrase used was “insurance fraud detection data mining”, the search was done in the time period between 1997 and 2013, it was conducted within published full text papers, and the result was a list of 566 papers.

III. Classification Framework on Data Mining and Insurance Fraud Detection

In this Section a classification framework is proposed for the available literature on the applications of data mining techniques to insurance fraud detection (IFD). The classification framework is based on previously published literature review papers [12], and existing knowledge on the nature of data mining research [13].

Our proposed classification framework for insurance fraud divides the insurance into three types automobile insurance, crop insurance and healthcare insurance. The data mining techniques used in insurance fraud detection are classified into six data mining application classes of classification, clustering, prediction, outlier detection, regression, and visualization. We provide a brief description of the six data mining application classes.

A. Classification:

Classification is defined as the act or process of putting things into groups based on ways that they are alike. In data mining classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known) [14]. Zhang and Zhou [15] state that classification and prediction is the process of identifying a set of common features and models that describe and distinguish data classes or concepts. Common classification techniques include neural networks, the naïve Bayes technique, decision trees and support vector machines. Such classification tasks are used in the detection of credit card, healthcare and automobile insurance, and corporate fraud, among other types of fraud, and classification is one of the most common learning models in the application of data mining in FFD.

B. Clustering:

Clustering is used to divide objects into conceptually meaningful groups (clusters), with the objects in a group being similar to one another but very dissimilar to the objects in other groups. Clustering is also known as data segmentation or partitioning and is regarded as a variant of unsupervised classification [14]. According to Yue et al. [16], "clustering analysis concerns the problem of decomposing or partitioning a data set (usually multivariate) into groups so that the points in one group are similar to each other and are as different as possible from the points in other groups." Further, Zhang and Zhou [15] argue that each cluster is a collection of data objects which are similar to one another within the same cluster but dissimilar to those in other clusters. The most common clustering techniques are the K-nearest neighbor, the Naïve Bayes technique and self-organizing map techniques.

C. Prediction:

Prediction is a statement about the way things will happen in the future based on experience or knowledge. P*rediction* is a statement that some outcome is expected. Prediction estimates numeric and ordered future values based on the patterns of a data set [13]. Han and Kamber [14] note that, for prediction, the attribute for which the values are being predicted is continuous-valued (ordered) rather than categorical (discrete-valued and unordered). This attribute can be referred to simply as the predicted attribute. Neural networks and logistic model prediction are the most commonly used prediction techniques.

D. Outlier detection:

Often there exist data objects that do not comply with the general behavior or model of the data. Such data objects, which are grossly different from or inconsistent with the remaining set of data, are called outliers. Many data mining algorithms try to minimize the influence of outliers or eliminate them all together. This, however, could result in the loss of important hidden information. In other words, the outliers may be of particular interest, such as in the case of fraud detection, where outliers may indicate fraudulent activity. Thus, outlier detection and analysis is an interesting data mining task, referred to as outlier mining. Outlier detection is employed to measure the "distance" between data objects to detect those objects that are grossly different from or inconsistent with the remaining data set [14] "Data that appear to have different characteristics than the rest of the population are called outliers" [17], p. 521]. Yamanishi et al. [18] point out that the problem of outlier/anomaly detection is one of the most fundamental issues in data mining. A commonly used technique in outlier detection is the discounting learning algorithm.

E. Regression:

Regression is a defensive reaction to some unaccepted impulses. It is statistical methodology used to reveal the relationship between one or more independent variables and a dependent variable [14]. Many papers have used logistic regression. The regression technique is typically undertaken using such mathematical methods as logistic regression and linear regression, and it is used in the detection of different types of fraud detection.

F. Visualization:

Data visualization is the creation and study of the visual representation of data, meaning, information that has been abstracted in some graphical form, including attributes or variables for the units of information. Visualization refers to the easily understandable presentation of the complex patterns or relationships uncovered in the data mining process [14]. Eick and Fyock [19] report that researchers at Bell and AT&T Laboratories have developed the pattern detection capability of the human visual system by building a suite of tools and applications that flexibly encode data using color, position, size and other visual characterist.

Type of fraud detection	Data mining application class	Data mining technique	Reference
Automobile insurance	Classification	Naïve Bayes	[20]
		Neural network, naïve Bayesian, decision trees	[21]
		Logistic model	[22]
		Neural networks, support vector machine, K-nearest neighbor, Naïve Bayes, Bayesian belief network, decision trees, Logistic model	[23]
		support vector machine, Genetic programming	[24]
		Decision tree , Naive Bayes tree , SVM-RFE (recursive feature elimination), SVM	[25]
		Consolidated Trees, decision trees	[26]
		Naïve Bayesian, Decision Tree	[27]
		Neural network, ensemble neural network	[28]
		Principal component analysis of RIDIT(PRIDIT)	[29]
		Logistic model	[30]
		Logistic model	[31]
		Logistic model	[32]
		Fuzzy logic	[33]
		Bayesian belief network, Logistic model	[34]
		Self-organizing map	[35]
		Fuzzy DEMATEL, Intuitionist fuzzy number, ELECTRE-TRI	[36]
		Gradient Boosting	[37]
	Prediction	Evolutionary algorithms, Cultural algorithms	[38]
		Social network analysis, Iterative Assessment Algorithm (IAA)	[39]
		Logistic model	[40]
	Regression	Probit model	[41]
		Logistic model	[42]
		Probit model	[43]
		Logistic model	[44]
Crop insurance		Yield-switching model	[45]
		Logistic model, probit model	[46]
Health care insurance	Classification	support vector machine, Gaussian (nonlinear), Linear kernels	[47]
		Naïve Bayes (NB), decision tree, Multiple Criteria Linear Programming (MCLP)	[48]
		Polymorphous (M-of-N) logic	[49]
		Self-organizing map	[50]
		Association rule	[51]
		decision tree	[52]
		support vector machine	[53]
	Clustering	SAS EM, CLUTO	[54]
		Nonnegative matrix factorization	[55]
		Regression analysis, Distances analysis	[56]
		Decision trees	[57]
	Outlier detection	Distance analysis, density estimation	[58]
		Risk	[59]
		R&DB-algorithm, RB-resolution algorithm	[60]
		Discounting learning algorithm	[18]
	Prediction	social network analysis, temporal analysis, higher order feature construction	[61]
	Visualization	Visualization	[62]

Table 1. Techniques used in insurance fraud detection

IV. Analysis of Insurance Fraud Detection

This paper provides a state-of-the-art review of the applications of data mining to insurance fraud detection. For the classification of insurance fraud, we divide the papers among the three types of insurance: automobile insurance (AI), crop insurance (CI) and healthcare insurance (HI). The second step is the classification of the 44 papers as given in Table 1. Table 1 lists the papers according to type of insurance fraud detection and the data mining application classes and techniques used with reference to the problems addressed. Some of the selected applications in the review address more than one IFD problem, and thus we categorized these applications by the dominant problem addressed. The third step makes more categorization using a set of algorithmic approaches (e.g., neural networks). Table 2 shows the classification of the 44 papers according to data application classes. Table 2 illustrates that most of the papers are on automobile insurance fraud (57%) than healthcare insurance fraud (39%). It also shows that classification is the most frequently used data mining application class 57% (41% + 16%) of the total (25 of the 44 papers), and that visualization is the least common, accounting for only 2.0% each (1 out of 44 each). Given that visualization is a significant method of fraud detection, which has characteristics that confer comparative advantages over other techniques, more attention should be paid to it in future research. To determine the main algorithms used for IFD, we present a simple analysis of IFD and the data mining techniques identified in the papers in table 3. Table 3 shows forty four data mining techniques used for insurance fraud detection. The most often used techniques are logistic models, Decision tree, the Na ïve Bayes, and support vector machine, all of which fall into the "classification" class. The logistic models are the most popular, being used in 10 of 44 of the papers reviewed, followed by Decision tree, used in 8 of the 44 papers, and then the Na ïve Bayes used 6 of the 44 papers and support vector machine used in 5 of the 44 papers.

Table 2. Distribution of papers by data mining application classes.

	Automobile insurance	AI	CI	HCI	Total
1	Logistic model	9	1		10
2	Decision Tree	5		3	8
3	Na ïve Bayes	5		1	6
4	support vector machine	3		2	5
5	Probit model	2	1		3
6	Bayesian belief network	2			2
7	distance analysis			2	2
8	Self-organizing map	1		1	2
9	Social network analysis	1		1	2
10	Neural network	2			2
11	Association rule			1	1
12	CLUTO			1	1
13	Consolidated Trees	1			1
14	Cultural algorithms	1			1
15	density estimation			1	1
16	Discounting learning algorithm			1	1
17	ELECTRE-TRI	1			1
18	Evolutionary algorithms	1			1
19	Fuzzy DEMATEL	1			1
20	Fuzzy logic	1			1
21	Gaussian (nonlinear)			1	1
22	Genetic programming	1			1
23	Gradient Boosting	1			1
24	higher order feature construction			1	1
26	Intuitionistic fuzzy number	1			1
27	Iterative Assessment Algorithm (IAA)	1			1
28	K-nearest neighbor	1			1
29	Linear kernels			1	1
30	Multiple Criteria Linear Programming			1	1
31	nonnegative matrix factorization			1	1
32	Polymorphous (M-of-N) logic			1	1
33	Principal component analysis of RIDIT	1			1
34	R&DB-algorithm			1	1
35	ensemble neural network	1			1
36	RB-resolution algorithm			1	1
37	regression analysis			1	1
38	risk			1	1
39	SAS EM			1	1
40	SVM (recursive feature elimination)	1			1
41	temporal analysis			1	1
42	text mining			1	1
43	Visualization			1	1
44	Yield-switching model		1		1

Table 4 shows the distribution of papers by type of insurance fraud and Publication year. It can be seen from this table that insurance fraud detection is an important research area and still is, since in 2011 there were 8 papers and 2012 there were 6 papers. The majority of these papers are in automobile insurance fraud detection, it is believed that it has to do with the data collection. It very difficult to collect data for insurance fraud detection in general but there is punch-data available for automobile fraud detection.

V. Conclusions

This paper proposes a classification framework for the application of data mining techniques to insurance fraud detection. This paper has explored almost all published insurance fraud detection papers in the five online databases that were searched. The search was done using several keywords to search online databases for papers published between 1997 and 2013.

The analysis results show that automobile insurance fraud in the most covered area of researched (57%). The data mining application class that was used in most of the papers is classification (57%). The most often used data mining techniques are logistic models, Decision tree, the Na ïve Bayes, and support vector machine, all of which fall into the "classification" class.

Table 3. Distribution of articles by data mining techniques

Type of fraud detection	Data mining application class	Number of papers	Percentage
Automobile insurance (AI)	Classification	18	41%
	Prediction	3	7%
	Regression	4	9%
	Total	**25**	**57%**
Crop insurance (CI)	Regression	2	5%
	Total	**2**	**5%**
Health care insurance (HCI)	Classification	7	16%
	Clustering	4	9%
	Outlier Detection	4	9%
	Prediction	1	2%
	Visualization	1	2%
	Total	**17**	**39%**
Total		**44**	

Table 4. Distribution of papers by publication year

Publication year	AI	CI	HCI	Total
1997	1		1	2
1998	2			2
1999	1			1
2000	1		1	2
2001			1	1
2002	6		1	7
2003				0
2004	1			1
2005	3	1		4
2006		1	2	3
2007	2		1	3
2008	1			1
2009				0
2010			1	1
2011	4		4	8
2012	3		3	6
2013			2	2
Total	**25**	**2**	**17**	**44**

References

[1] J. Pearsall, *The concise Oxford dictionary-/ed. by Judy Pearsall*: Oxford [etc.]: Oxford University Press, 1999.

[2] C. hua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010.

[3] J.-H. Wang, Y.-L. Liao, T.-m. Tsai, and G. Hung, "Technology-based Financial Frauds in Taiwan: Issues and Approaches," in *SMC*, 2006, pp. 1120-1124.

[4] "BBC News," http://news.bbc.co.uk/1/hi/business/6636005.stm, 2007.

[5] C. Elkan, "Magical thinking in data mining: lessons from CoIL challenge 2000," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 426-431.

[6] N. Lavrač, H. Motoda, T. Fawcett, R. Holte, P. Langley, and P. Adriaans, "Introduction: Lessons learned from data mining applications and collaborative problem solving," *Machine learning*, vol. 57, pp. 13-34, 2004.

[7] I. Bose and R. K. Mahapatra, "Business data mining—a machine learning perspective," *Information & management*, vol. 39, pp. 211-225, 2001.

[8] E. Turban, R. Sharda, D. Delen, and T. Efraim, *Decision support and business intelligence systems*: Pearson Education India, 2007.

[9] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge discovery in databases: An overview," *AI magazine*, vol. 13, p. 57, 1992.

[10] Y. Kou, C.-T. Lu, S. Sirwongwattana, and Y.-P. Huang, "Survey of fraud detection techniques," in *Networking, sensing and control, 2004 IEEE international conference on*, 2004, pp. 749-754.

[11] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010.

[12] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, pp. 559-569, 2011.

[13] S. R. Ahmed, "Applications of data mining in retail business," in *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*, 2004, pp. 455-459.

[14] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*: Morgan kaufmann, 2006.

[15] D. Zhang and L. Zhou, "Discovering golden nuggets: data mining in financial application," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 34, pp. 513-522, 2004.

[16] D. Yue, X. Wu, Y. Wang, Y. Li, and C.-H. Chu, "A review of data mining-based financial fraud detection research," in *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, 2007, pp. 5519-5522.

[17] M. Agyemang, K. Barker, and R. Alhajj, "A comprehensive survey of numeric and symbolic outlier mining techniques," *Intelligent Data Analysis*, vol. 10, pp. 521-538, 2006.

[18] K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000, pp. 320-324.

[19] S. G. Eick and D. E. Fyock, "Visualizing corporate data," *Potentials, IEEE*, vol. 15, pp. 6-11, 1996.

[20] S. Viaene, R. A. Derrig, and G. Dedene, "A case study of applying boosting Naive Bayes to claim fraud diagnosis," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, pp. 612-620, 2004.

[21] S. Viaene, G. Dedene, and R. A. Derrig, "Auto claim fraud detection using Bayesian learning neural networks," *Expert Systems with Applications*, vol. 29, pp. 653-666, 2005.

[22] M. Artís, M. Ayuso, and M. Guillén, "Modelling different types of automobile insurance fraud behaviour in the Spanish market," *Insurance: Mathematics and Economics*, vol. 24, pp. 67-81, 1999.

[23] S. Viaene, R. A. Derrig, B. Baesens, and G. Dedene, "A Comparison of State - of - the - Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection," *Journal of Risk and Insurance*, vol. 69, pp. 373-421, 2002.

[24] M. Vasu and V. Ravi, "A hybrid under-sampling approach for mining unbalanced datasets: applications to banking and insurance," *International Journal of Data Mining, Modelling and Management*, vol. 3, pp. 75-105, 2011.

[25] M. Farquad, V. Ravi, and S. B. Raju, "Analytical CRM in banking and finance using SVM: a modified active learning–based rule extraction approach," *International Journal of Electronic Customer Relationship Management*, vol. 6, pp. 48-73, 2012.

[26] J. M. Pérez, J. Muguerza, O. Arbelaitz, I. Gurrutxaga, and J. I. Martín, "Consolidated tree classifier learning in a car insurance fraud detection domain with class imbalance," in *Pattern Recognition and Data Mining*, ed: Springer, 2005, pp. 381-389.

[27] R. Bhowmik, "Detecting auto insurance fraud by data mining techniques," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, pp. 156-162, 2011.

[28] W. Xu, S. Wang, D. Zhang, and B. Yang, "Random Rough Subspace Based Neural Network Ensemble for Insurance Fraud Detection," in *Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on*, 2011, pp. 1276-1280.

[29] P. L. Brockett, R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert, "Fraud classification using principal component analysis of RIDITs," *Journal of Risk and Insurance*, vol. 69, pp. 341-371, 2002.

[30] S. Viaene, M. Ayuso, M. Guillen, D. Van Gheel, and G. Dedene, "Strategies for detecting fraudulent claims in the automobile insurance industry,"

European Journal of Operational Research, vol. 176, pp. 565-583, 2007.

[31] M. Artís, M. Ayuso, and M. Guillén, "Detection of automobile insurance fraud with discrete choice models and misclassified claims," *Journal of Risk and Insurance*, vol. 69, pp. 325-340, 2002.

[32] S. B. Caudill, M. Ayuso, and M. Guillen, "Fraud detection using a multinomial logit model with missing information," *Journal of Risk and Insurance*, vol. 72, pp. 539-550, 2005.

[33] J. Pathak, N. Vidyarthi, and S. L. Summers, "A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims," *Managerial Auditing Journal*, vol. 20, pp. 632-644, 2005.

[34] L. Bermúdez, J. Pérez, M. Ayuso, E. Gómez, and F. Vázquez, "A Bayesian dichotomous model with asymmetric link for fraud in insurance," *Insurance: Mathematics and Economics*, vol. 42, pp. 779-786, 2008.

[35] P. L. Brockett, X. Xia, and R. A. Derrig, "Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud," *Journal of Risk and Insurance*, pp. 245-274, 1998.

[36] E. Taghiloo, A. Azar, E. Nasiri, and H. Ghaedrahmati, "A New Model for Insurance Fraud Detection in Car Accidents Using a Combined Fuzzy DEMATEL and ELECTRE-TRI Approach."

[37] L. Guelman, "Gradient boosting trees for auto insurance loss cost modeling and prediction," *Expert Systems with Applications*, vol. 39, pp. 3659-3667, 2012.

[38] M. Sternberg and R. G. Reynolds, "Using cultural algorithms to support re-engineering of rule-based expert systems in dynamic performance environments: a case study in fraud detection," *Evolutionary Computation, IEEE Transactions on*, vol. 1, pp. 225-243, 1997.

[39] L. Šubelj, Š. Furlan, and M. Bajec, "An expert system for detecting automobile insurance fraud using social network analysis," *Expert Systems with Applications*, vol. 38, pp. 1039-1052, 2011.

[40] S. Tennyson and P. Salsas‐Forn, "Claims auditing in automobile insurance: fraud detection and deterrence objectives," *Journal of Risk and Insurance*, vol. 69, pp. 289-308, 2002.

[41] E. B. Belhadji, G. Dionne, and F. Tarkhani, "A model for the detection of insurance fraud," *Geneva Papers on Risk and Insurance. Issues and Practice*, pp. 517-538, 2000.

[42] K. J. Crocker and S. Tennyson, "Insurance fraud and optimal claims settlement strategies," *JL & Econ.*, vol. 45, p. 469, 2002.

[43] J. Pinquet, M. Ayuso, and M. Guillen, "Selection bias and auditing policies for insurance claims," *Journal of Risk and Insurance*, vol. 74, pp. 425-440, 2007.

[44] H. I. Weisberg and R. A. Derrig, "Quantitative methods for detecting fraudulent automobile bodily injury claims," *Risques*, vol. 35, pp. 75-101, 1998.

[45] J. A. Atwood, J. F. Robison-Cox, and S. Shaik, "Estimating the prevalence and cost of yield-switching fraud in the federal crop insurance program," *American journal of agricultural economics*, vol. 88, pp. 365-381, 2006.

[46] Y. Jin, R. M. Rejesus*, and B. B. Little, "Binary choice models for rare events data: a crop insurance fraud application," *Applied Economics*, vol. 37, pp. 841-848, 2005.

[47] M. Kirlidog and C. Asuk, "A Fraud Detection Approach with Data Mining in Health Insurance," *Procedia-Social and Behavioral Sciences*, vol. 62, pp. 989-994, 2012.

[48] Y. Peng, G. Kou, A. Sabatka, J. Matza, Z. Chen, D. Khazanchi, and Y. Shi, "Application of classification methods to individual disability income insurance fraud detection," in *Computational Science–ICCS 2007*, ed: Springer, 2007, pp. 852-858.

[49] J. A. Major and D. R. Riedinger, "EFD: A Hybrid Knowledge/Statistical‐Based System for the Detection of Fraud," *Journal of Risk and Insurance*, vol. 69, pp. 309-324, 2002.

[50] H. He, J. Wang, W. Graco, and S. Hawkins, "Application of neural networks to detection of medical fraud," *Expert Systems with Applications*, vol. 13, pp. 329-336, 1997.

[51] W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Systems with Applications*, vol. 31, pp. 56-68, 2006.

[52] K. C. Lin, C. L. Yeh, and S. Y. Huang, "Use of Data Mining Techniques to Detect Medical Fraud in Health Insurance," *Applied Mechanics and Materials*, vol. 284, pp. 1574-1578, 2013.

[53] C. Francis, N. Pepper, and H. Strong, "Using support vector machines to detect medical fraud and abuse," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 8291-8294.

[54] Y. Peng, G. Kou, A. Sabatka, Z. Chen, D. Khazanchi, and Y. Shi, "Application of Clustering Methods to Health Insurance Fraud Detection," in *Service Systems and Service Management, 2006 International Conference on*, 2006, pp. 116-120.

[55] S. Zhu, Y. Wang, and Y. Wu, "Health care fraud detection using nonnegative matrix factorization," in *Computer Science & Education (ICCSE), 2011 6th International Conference on*, 2011, pp. 499-503.

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

[57] H. Shin, H. Park, J. Lee, and W. C. Jhee, "A scoring model to detect abusive billing patterns in health insurance claims," *Expert Systems with Applications*, vol. 39, pp. 7441-7450, 2012.

[58] R. M. Konijn and W. Kowalczyk, "Finding fraud in health insurance data with two-layer outlier detection approach," in *Data Warehousing and Knowledge Discovery*, ed: Springer, 2011, pp. 394-405.

[59] K. D. Aral, H. A. Güvenir, İ. Sabuncuoğlu, and A. R. Akar, "A prescription fraud detection model," *Computer methods and programs in biomedicine*, vol. 106, pp. 37-46, 2012.

[60] W. Xiaoyun and L. Danyue, "Hybrid outlier mining algorithm based evaluation of client moral risk in insurance company," in *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, 2010, pp. 585-589.

[61] V. Chandola, S. R. Sukumar, and J. C. Schryver, "Knowledge discovery from massive healthcare claims data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1312-1320.

[62] L. Sokol, B. Garcia, M. West, J. Rodriguez, and K. Johnson, "Precursory steps to mining HCFA health care claims," in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, 2001, p. 10 pp.