IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-11 E-ISSN: 2347-2693

Applications of Data Mining in Fraud Detection

Fuzail Misarwala, KausarMukadam, and Kiran Bhowmick.

Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, India.

www.ijcseonline.org

Received: Oct/22/2015 Revised: Nov /05/2015 Accepted: Nov/20/2015 Published: Nov/30/2015

***Abstract*-**It comes as no surprise to learn that from an economic standpoint, fraud continues to be a growing concern for organisations of all sizes, across all regions and in virtually every sector. A 2014 survey shows that 5% of the losses at an organization can be attributed to fraud, which applied to the Gross World Producttranslates to a projected global fraud revenue loss of nearly $3.7 trillion. [1] Due to ever increasing volume of data that needs to be analysed in order to detect these frauds, data mining methods and techniques are being used with increasing frequency in this domain. This paper is aimed at providing an expansive literature review of journal articles produced between 2008 and 2015 to demonstrate the extensive research that has been carried out in selected domains and also to highlight the gaps between industry need and research in the particular areas. We have classified the research papers based on the data mining technique used, the type of fraud targeted, year of publishing, etc. and analysed the results.

***Keywords*—**Data Mining; Fraud; Fraud Detection; Classification; Support Vector Machine; Computer Intrusion

I. INTRODUCTION

This paper proposes a framework to classify and review the most recent research in the field of fraud detection through the use of data mining.Considering the recent increase in fraudulent activities, and subsequent efforts to tackle them, summarization of research within this period will aid in further accretion of knowledge under this domain.

Any unlawful or unfair gain by deliberate deception is termed in law, as fraud which is both a civil wrong and a criminal wrong There are multiple types of frauds such as Financial frauds, subdividing into credit card frauds, money laundering, insurance fraud, bank frauds etc. as well as Telecommunications frauds, Medical and Scientific frauds to name a few.

Generally, data mining (sometimes called data or knowledge discovery) is the process of the analysis of data from variousoutlooks and summarizing it into useful information to increase revenue, cut costs, or both. It allows users to analyse data from many different dimensions or angles, categorize it, and effectively summarize the relationships identified. It is the process of finding correlations or associations among a multitude of fields in large relational databases[2].

In recent years, data mining methods have been increasingly used along with numeric data to develop fraud detection systems or predictive models. All the information that can be associated with a record under survey, such as an insurance claim, credit application, or a purchase, is analysedand used to improve accuracy of the detection system.

II. METHODICAL FRAMEWORK FOR RESEARCH

The goal of the paper is to research the applications of data mining in fraud detection during the period between 2008 and 2015. Using this survey we have tried to determine which aspect of fraud detection has garnered the most interest and which fields still lack research. Also, we have tried to conclude the most widely used data mining techniques in fraud detection practices. Further, in research methodology, the specific criteria for selecting and including papers in this research have been specified. Three journal databases, including IEEE Transactions, Science Direct, and SPRINGER, were searched for against the keywords “fraud”, “fraud detection” and “data mining” through the use of logical operators (AND and OR) to yield appropriate results. The “fraud” and “fraud detection” descriptors cover all the various categories of fraud including (but not limited to) insurance fraud, telecommunications fraud, bank fraud, and credit card fraud. The absence of specific keywords such as “credit card fraud”, “insurance fraud”, etc. allowed the data set to remain unbiased towards any particular category. This search yielded approximately 800 articles, from which around 90 were selected on basis of relevance. These articles were further analysed by the authors to ensure that only the most relevant articles, i.e. articles researching applications of data mining to fraud detection, were chosen. Other criteria applied during selection included year of publishing, as only articles during the years 2008-2015 were chosen, and availability. Unpublished working papers, textbook extracts and conference papers were excluded as

their full texts are not currently available. In the end only full text, published articles were chosen for the study.

III. DATA MINING TECHNIQUES

Data mining involves eight common classes of tasks:

Association: Association (also called relation technique) discovers patterns based on a relationship between items in the same transaction. It finds rules existing in the database that fulfil some minimum support and confidence constraints [5]. It can be used in market basket analysis for identifying a set of products that customers frequently purchase together and also in systems such as intrusion detection, heterogeneous genome data, mining remotely sensed images/data and product assortment decisions [3].

Classification: Classification (or supervised learning) is a classic data mining technique based on machine learning that involves learning a function that maps (or classifies) a data item into one of the previously defined classes[4]. It makes use of mathematical techniques such as linear programming, decision trees, neural networks and statistics.

Clustering: Clustering (or unsupervised learning) is a data mining technique that makes meaningful or useful cluster of objects based on similar characteristics. Clustering is predominantly done to analyse and use hidden information present in groups or to find conceptually meaningful groups with collective characteristics [7].

Regression: Regression is a data mining (or machine learning) technique utilized to fit an equation to a dataset. It is a data mining function that can be used to predict a number such as profit, mortgage, temperature, or distance, For example, a regression model can predict the value or price of a house based on size, location, parking space, etc.

Summarization: Summarization provides compact representation of the data set, and can be used in visualization and report generation. It reduces the size and complexity of enormous multidimensional datasets to more manageable proportions through sophisticated methods such as derivation of summary rules, discovery of functional relationships between variables and multivariate visualization technique[4].

Anomaly (Outlier) Detection: In data mining, anomaly detection (or outlier detection) is the identification of items, events or observations that do not conform to an expected pattern or other items in a dataset. Typically these anomalous items will translate to some kind of problem such as bank fraud, a structural defect, medical problems or errors in a text.

Visualization: Data visualization can be described as an effort made to understand the importance of data by placing it in a visual context. Through visualization software and techniques, patterns and trends that may go undetected in text-based data can be recognized. Visualization is aimed at applying perceptual ability to large data sets in computer systems[5].

Prediction: Prediction in data mining is the analysis of past and present facts to predict future events. Identifiable risks and opportunities are captured through the analysis of existing data, by identifying related factors and the relationships between them. Predictive analysis is used in multiple scientific and business analytic fields, as well as fraud detection.

IV. FRAUD CLASSIFICATION

The framework for fraud classification is depicted in figure 1.

A. Financial Fraud

Financial fraud can be described as theft or larceny through which a person/entity takes property or money, or uses them in an illegal manner, with intent to gain a benefit from it[6]. Due to the complex economy prevalent today, such fraud and crimes can take many forms such as bank fraud[7][8][9][10], securities and commodities fraud[11], occupational (or internal) fraud [12][13], taxpayer fraud[14], money laundering[15], financial statement fraud[16][9]and advanced fee fraud[17]. For the purpose of classification, in this paper, financial fraud is divided into:

a) Bank Fraud

According to Connell University Law School (CULS), bank fraud is defined as "whoever knowingly executes, or attempts to execute, a scheme or artifice (1) to defraud a financial institution; or (2) to obtain any of the moneys, funds, credits, assets, securities, or other property owned by, or under the custody or control of, a financial institution, by means of false or fraudulent pretences, representations, or promises"[18].

For the purpose of this study, bank fraud includes online banking fraud[19], credit card fraud, money laundering, and mortgage fraud. Credit card fraud can be described as identity theft that involves an unauthorized use of another person's credit card information for the purpose of charging purchases or withdrawing funds from it [20], whereas money laundering is defined as the processing, i.e. disguising, of criminally acquired proceeds to hide their illegal origins and transform them into outwardly legal transactions.

Emanuel MinedaCarneiro et al. [21] conducted research experiments that used 645,538 real internet credit card transactions, out of which 37,359 were fraudulent. Through the use of Multilayer Perceptron Artificial Neural Networks and Cluster Analysis (Iterative Naïve Bayesian Inference

Agglomerative Clustering algorithm), credit-card fraud was detected.

Traditional research in money laundering concentrates mainly on k-means clustering technique due to the previously promising results. Recently, research on use of Expectation Maximization (EM) for Anti-Money Laundering (AML) has also developed. Zhiyuan Chen et al.[15]have explored and exploited the advantages of EM for AML with promising results.

b) Securities and Commodities Fraud:

Investment fraud, a common type of securities and commodities fraud, involves schemes (also called high yield investment fraud) that involve illegal sale or purported sale of financial instruments. These include Ponzi schemes, Pyramid schemes, Prime bank investment fraud/trading program fraud, Advance fee fraud[17]and Broker embezzlement, among others.

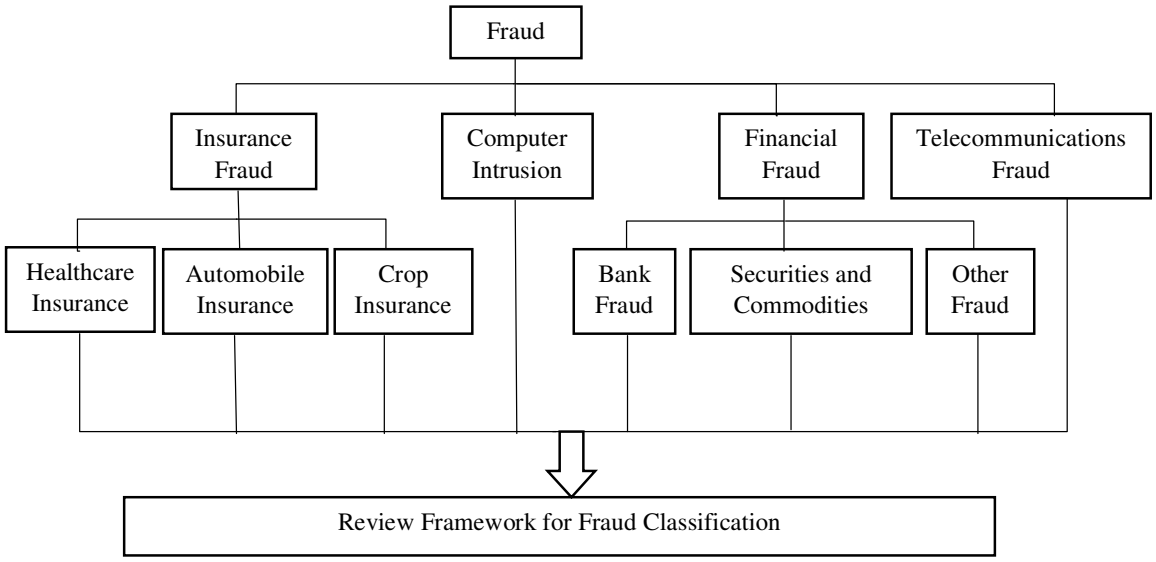

Figure 1. Review Framework

Commodities fraud is the sale or purported sale of a commodity i.e. raw materials or semi-finished goods that are sold on an exchange such as gold or coffee, through illegal means.

KooshaGolmohammadi and Osmar R. Zaiane[11] presented a literature survey of 205 papers and references in relation to securities fraud and discussed the use of Pattern Recognition, Outlier Detection, Rule Induction, Social Network Analysis and Visualization in this domain. Several challenges faced in securities fraud detection such as High Frequency Trading (HFT), unlabelled data, massive data sets and variations in data forms, were also highlighted.

c) Other Related Types of Fraud

Tax Payer Fraud, Occupational Fraud and Financial Statement fraud are some of the frauds that are included in this category.

Occupational Fraud can be defined as "the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets" [1]. Various models like use of data mining along with Process Aware Systems [13] and Information Visualization [12] have been proposed to tackle Occupational Fraud.

Treadway Commission[22] defines Financial Statement Fraud as "Any intentional act or omission that results in materially misleading financial statements." In our review different methods for Financial Statement Fraud detection such as Multilayer Feed Forward Neural Network (MLFF), Logistic Regression (LR), Genetic Programming (GP), Group Method of Data Handling (GMDH), Support Vector Machines (SVM), and Probabilistic Neural Network (PNN)[23] in addition to Distance Weighted Discrimination (DWD) [24] have been included.

B. Computer Intrusion

Compromising a computer system by breaking the security or causing it to enter into an insecure or unstable state is known as computer intrusion. This act of intruding—or obtaining unauthorized access to a system—which leaves traces that can be discovered through an intrusion detection system. Two types of intrusion detection systems are:

Misuse detection systems that match computer activities with previously known attacks in their database, and Anomaly detection systems that by learning normal activity, detect any activities that deviate from normal patterns

Manoj Kumar et al.[25]proposed a method for intrusion detection in cloud computing, using an outlier detection concept, Density Based Outlier Detection (DenOD),which is an unsupervised technique for detection of network fraud without previous knowledge of attacks. It is implemented using Intrusion Detection in Cloud Computing (IDCC) Framework through the use of Cloud nodes and an IDS.

An XCS (Accuracy based Learning Classifier System) based network intrusion detection model is proposed by Mohammad Behdad et al. [26]. The performance of this XCSR has been tested on various parameters such as different levels of output bias, adaptability in incomplete domains, effect of noise on performance, effect of concept drift and targeted bias drift , and finally was applied to real-world data (KDD Cup 1999 Intrusion).

C. Insurance Fraud

Insurance frauds involve an insurance company, agent or other person being deceived by individuals in an attempt to achieve monetary gains to which they are not entitled. Insurance fraud is said to have happened when someone has put false information on an insurance application and when misleading or incorrect information is given or crucial information is omitted in an insurance claim or transaction. We consider multiple areas of insurance namely Healthcare insurance, automobile insurance, and corporate insurance to analyse research.

A convincing technique constructed and tested by Han Tao et. al[27] is a fuzzy support vector machine model, which is based on traditional SVM with dual membership for the identification of insurance fraud. Their empirical results show an impressive 90.73% precision on a recall of 91.31% when using dual membership fuzzy support vector machine model for fraud identification, which is higher compared to other models tested by them.

A slightly different approach was taken by MelihKirlidoget. al[28], who performed anomaly detection on an Oracle system that used SVM algorithm. Their idea was to try and detect anomalous claims by calculating the probability of a claim being fraudulent where any record that had a probability rating of over 50% was investigated against multiple criteria to determine whether the record is really fraudulent, which may lead to the revelation of some unknown patterns in malicious insurance claims.

D. Telecommunications Fraud

Joe Ariganello, in his book, defines telecommunications fraud as “the theft of telecommunication services or the use of this service to commit other forms of fraud”[29]. Rapidly changing technology has brought in increasingly sophisticated and complex methods for fraudsters to infiltrate technology alongside the complex innovations. While detection of such fraudulent activities may be challenging, they are not ignored as consistent research in this field keeps a high degree of fraudsters at bay.

One such research is by Anna Leontjeva et al [30] who use a hypergraph classifier (along with distance-based approach and kernel-based approach) for detection of possible telecommunications frauds. They present empirical results with a recall of 16% and a high precision rate on data acquired from an internet call company.

Another paper [31] presents a framework for fraud detection targeting VoIP or network OSS/BSS network vulnerabilities through the use of an ontology model. They tackle the difficulties encountered in VoIP fraud detection by reviewing the SIP security issues that probably cause frauds. The paper presents a SIP based framework which is based on user profiling, hence creating an effective intrusion detection system.

Other aspects of fraud, such as subscription fraud[32] and occurrence of fraudulent calls from mobile phone users [33] have also been explored.

V. ANALYSIS

This paper provides a thorough literature review of the application of data mining in different types of fraud detection. A complete summary and organization of literature is given in Table.1.

For the purpose of classification, fraud has been divided into four broad categories- financial fraud, telecommunications fraud, computer intrusion and insurance fraud. Financial fraud is further divided into bank fraud, securities and commodities fraud and other types of related fraud which includes financial statement fraud, tax payer fraud and occupational fraud, whereas Insurance fraud is further classified into health insurance fraud, crop insurance fraud and automobile insurance fraud. As the 47 papers covered in this survey did not include crop insurance fraud, it has not been mentioned in table 1.

In terms of data mining, the articles have been divided on the basis of eight data mining application classes

which are association, classification, clustering, visualization, prediction, outlier detection, summarization, regression. Again, as no papers were included that used association and summarization, these classes have not been included in the table. The articles have then been further

categorized on the basis of the actual data mining algorithm or technique used, such as Naïve Bayes, Support Vector Machine, Self-Organizing Map, etc. A detailed analysis of the 47 articles surveyed is given in the next section.

A. Distribution of articled by fraud type

Judging by the distribution of articles in Table. 1, it can be clearly noticed that that most amount of research has been conducted in the field of financial fraud. From the

Table 1: Distribution of articles by type of fraud.

Fraud	Sub-Category	Data mining application class	Data mining techniques	Referenc es
Insurance Fraud	Health Insurance Fraud	Classification	Support Vector Machine	[34]
		Clustering	Evolving Clustering Method	[34]
		Outlier Detection	One-Class Support Vector Machine	[28]
	Automobile Insurance Fraud	Classification	Logistic Model, Bayesian Belief Network	[35]
			Support Vector Machine	[27]
Telecommunication Fraud		Classification	Bayesian Belief Network	[31]
			Support Vector Machine	[36]
			Support Vector Machine	[37]
			Support Vector Machine, K-nearest neighbour	[30]
			Feed Forward Neural Network	[38]
			Support Vector Machine, Neural Networks, Decision Tree	[32]
			Not Specified	[39]
		Clustering	Self-Organizing Map	[40]
			Not Specified	[39]
			K-means, Self-Organizing Maps	[32]
			Hierarchical Agglomerative Clustering	[41]
		Outlier Detection	One-Class Support Vector Machine	[33]
		Prediction	Not Specified	[39]
			Decision tree	[42]
		Regression	Logistic regression	[30]
		Visualization	Self-Organizing Map	[43]
			Self-Organizing Map	[40]
Computer Intrusion		Classification	Learning Classifier System (Accuracy Based)	[26]
		Outlier Detection	Density Based Outlier Detection	[25]
		Predictive	CART	[44]
		Visualization	Self-Organizing Map	[43]
Financial Fraud	Securities and Commodities Fraud	Classification	Random Forest, Support Vector Machine	[17]
	Bank Fraud	Classification	Aggregation -Random Forest	[45]
			Genetic Algorithm	[46]
			Bayesian Classification	[8]
			Neural Network, Contrast Pattern Mining, Decision Forest.	[19]
			Very Fast Decision Trees	[47]
			Neural Networks	[21]
			Genetic Algorithm, Inductive Learning	[48]
			BOAT Algorithm	[49]
			Artificial Neural Network, Multi-Layer Perceptron, Decision Tree	[50]
		Clustering	Self-Organizing Map	[7]
			K-Means Algorithm	[51]
			Expectation Maximizing Algorithm	[15]
			Cluster Analysis	[21]
			K-Means Clustering	[49]
			Gaussian Mixture Model	[52]
		Outlier Detection	Similar Coefficient Sum	[53]
			Distance-Sum	[54]
			Bagging Ensemble Classifier	[55]
		Summarization	Social Media Crowdsourcing	[56]
		Visualization	Self-Organizing Map	[43]
	Other Related Financial Fraud	Classification	Genetic Algorithm, Distance Weighted Discrimination	[24]
			Naïve Bayes Algorithm	[10]
			Neural Network, Genetic Programming, Support Vector Machine	[23]
			Bayesian Networks, Neural Networks, Decision trees	[14]
			Support Vector Machine	[9]
		Classification	Social Network Analysis	[57]
		Clustering	Self-Organizing Map, K-means	[58]

		Self-Organizing Map, Neural gas	[14]
		Hierarchical Clustering Algorithm	[13]
	Regression	Logical Regression	[23]
	Summarization	Social Media Crowdsourcing	[56]

47 articles reviewed in this paper, 61.7% belong to the domain of financial fraud detection. Within financial fraud detection, maximum research has been conducted in bank fraud (36.17%), with credit card fraud contributing almost 25.5% of the research. Along with credit card fraud, online-bank fraud and account frauds have also been investigated. Within financial fraud, the second most prominent area is other types of related fraud (21.2%) from which financial statement fraud (6.38%) has garnered the most attention. Securities and Commodities fraud have been researched far less in comparison, amounting to only 4.2% of 47 articles published. Telecommunication fraud detection contributes 21.3% of the total number of research articles, making it the second most researched sector (after financial fraud).This is followed by computer intrusion (8.5%) and insurance fraud (8.5%) –health insurance fraud (4.25%) and automobile insurance fraud (4.25%).

B. Distribution of articles by year

The distribution of articles by year and fraud type is specified in Table. 2.

In the past couple of years, research in computer intrusion and health insurance fraud detection has seen a rise and these fields will continue to develop. Bank fraud detection has received a significant amount of importance through all these 8 years. Being one of the most widespread and common frauds, research in this field will also continue to advance. In certain domains such as securities and commodities fraud, and automobile insurance, very little research has been conducted. This could be because of the lack of sufficient data sets or due to the sensitive nature of cases (in the case of securities fraud).

C. Distribution of articles by data mining application classes

Classification (59.57%) is the most applied data mining class, followed by clustering (29.78%). In a previous survey of literature on fraud [64] outlier detection and visualization had accounted for only 2% of the articles researched. In recent years, more attention has been paid to these areas, increasing the number of articles to 14.89 % (Outlier Detection) and 8.51% (Visualization).

D. Distribution of articles by data mining techniques or algorithm

As shown in Table. 3, 31 techniques have been applied to detect fraud, out of which the most frequently used are:

Support Vector Machine (13.63%): It is a supervised learning model that analyses data and can recognize patterns in regression analysis and classification. It builds a model based on training examples and assigns new examples to these models. It has been used in telecommunications fraud, bank fraud, other financial frauds and insurance fraud

Neural Network (10.6%): This technique imitates the functionalities of the human brain by using interconnected vertices. It can be used in clustering as well as classification. They have been applied to bank fraud and telecommunication fraud detection.

Decision tree (7.55%): Decision tree is a predictive support tool that uses possible outcomes of decisions through a tree-like model for mapping. Leaves represent class labels, branches represent outcomes and each internal node depicts a test. Decision trees have been used in financial and telecommunications fraud.

Self-Organizing Map (7.55%): Self-Organizing Maps are a type of artificial neural network that uses unsupervised learning. It maps the high dimensional space to map units and is used in visualization. Self-Organizing Maps have been used in financial and telecommunication fraud detection.

Table 2: Distribution of articles by year.

		2008	2009	2010	2011	2012	2013	2014	2015	Total
Insurance Fraud	Automobile Insurance	1				1				2
	Health Insurance					1			1	2
Financial Fraud	Bank Fraud	1	2	3	1	4	1	3	2	17
	Securities and Commodities				1	1				2

Other Related Fraud		2	1	1	1	3	2		10
Telecomm. Fraud	3	2	2	1	1			1	10
Computer Intrusion					2		2		4
Total	5	6	6	4	11	4	7	4	47

Table 3: Distribution of articles by data mining technique used.

Technique/Algorithm	Financial Fraud			Insurance Fraud		Computer Intrusion	Telecommunications Fraud	Total
	Bank Fraud	Securities and Commodities Fraud	Other types of related fraud	Health Insurance Fraud	Automobile Insurance Fraud			
Aggregation -Random Forest	1							1
Neural Network	3		2				2	7
Bagging Ensemble Classifier	1							1
Bayesian Belief Network					1		1	2
Bayesian Classification	1		1					2
BOAT Algorithm	1							1
CART						1		1
Cluster Analysis	1							1
Contrast Pattern Mining	1							1
Decision Tree	2		2				2	6
Density Based Outlier Detection						1		1
Distance Weighted Discrimination			1					1
Distance-Sum	1							1
Evolving Clustering Method				1				1
Expectation Maximizing Algorithm	1							1
Gaussian Mixture Model	1							1
Genetic Algorithm	2		2					4
Hierarchical Clustering Algorithm	1						1	2
Inductive Learning	1							1
K-Means Clustering	2		1				2	5
Learning Classifier System (Accuracy Based)						1		1
Logical Regression			1		1		1	3
Multi-Layer Perceptron	1							1
Naïve Bayes Algorithm			1					1
One-Class Support Vector Machine				1			1	2
Random Forest		1						1
Self-Organizing Map	1		2				3	6
Similar Coefficient Sum	1							1
Social Network Analysis			1					1
Support Vector Machine		1	2	1	1		4	9
Very Fast Decision Trees	1							1
Total	24	2	16	3	3	3	17	68

VI. CONCLUSION

Although we have tried to provide a thorough review of all relevant research in this field, like most other review, our survey had some limitations. The conclusions of our review and its limitations have been described below.

Our aim was to present a review that can inform both academicians and industry professionals about the current state of research in this field in terms of sectors targeted and methods used.

We classified the literature based on (i) Fraud type (ii) Year (iii) Data mining class, and (iv) Data mining technique. Out of the four categories of fraud that we investigated, financial fraud has attracted the most attention from researchers. Financial fraud is morelikely to be committed by offenders and affects businesses and organizations of all sizes, thus it is a matter of grave concern for most.May be this is why most research has been conducted in this domain. Also, organized data sets are more easily available in finance as compared to field such as intrusion detection and telecommunications. Within financialfraud, mortgage fraud and securities and commodities fraud are lacking in research as compared to other sub categories. Reasons for this could be insufficient research data or sensitivity of such frauds. Nevertheless, more research is required in this domain.

Due to fast paced technologies, computer intrusion has been on the rise. While recently research has progressed in this field, it still remains far behind the growing industrial need.

In this research, only 47 articles have been analysed due to constraints such as access to journals and articles. Further scope in this field would be to extend this research to articles published in languages other than English and also to cover various other English publications. Another advancement could be to analyse research over a greater period to map the changes in trends in detection of different types of fraud, so as to correctly determine the future of research in fraud detection.

VII. REFERENCES

[1] "Report to the Nations on Occupational Fraud and Abuse: 2014 Global Study," 2014. [Online]. Available: http://www.acfe.com/rttn/docs/2014-report-to-nations.pdf.

[2] B. Palace, *What is data mining?,* Anderson Graduate School of Management at UCLA, 1996.

[3] A. W.-C. F. Raymond Chi-Wing Wong, *Association Rule Mining and its Application to MPIS.*

[4] G. P.-S. a. P. S. Usama Fayyad, "From Data Mining to Knowledge Discovery in Databases," *American Association for Artificial Intelligence,* 1996.

[5] D. A. Keim, "Information Visualization and Data Mining," *IEEE Transactions on Visualization and Computer Graphics,* 2002.

[6] "FindLaw," [Online]. Available: http://criminal.findlaw.com/criminal-charges/fraud-financial-crimes.html.

[7] G. C. Y. B. N. Grozavu, "Unsupervised Learning for Analyzing the Dynamic Behavior of Online Banking Fraud," *International Conference on Data Mining Workshops IEEE,* 2013.

[8] D. C. Y. ,. W.-H. L. ,. C. W. Shing-Han Li, "Identifying the signs of fraudulent accounts using data mining techniques," *Computers in Human Behavior Elsevier,* 2012.

[9] Q. Deng, "Application of Support Vector Machine in the Detection of Fraudulent Financial Statements," *4th International Conference on Computer Science & Education,* 2009.

[10] Q. Deng, "Detection of Fraudulent Financial Statements Based on Naïve Bayes Classifier," *The 5th International Conference on Computer Science & Education,* 2010.

[11] O. R. Z. Koosha Golmohammadi, "Data Mining Applications for Fraud Detection in Securities Market," *European Intelligence and Security Informatics Conference IEEE,* 2012.

[12] A. A. S. A. S. Evmorfia N. Argyriou, "Occupational Fraud Detection Through Visualization," *IEEE,* 2013.

[13] H. R. S. Shahla Mardani, "A New Method for Occupational Fraud Detection in Process Aware Information Systems," *International ISC Conference on Information Security and Cryptology IEEE,* 2013.

[14] J. D. V. Pamela Castellón González, "Characterization and detection of taxpayers with false invoices using data mining techniques," *Expert Systems with Applications- Elsevier,* 2013.

[15] L. D. V. K. N. E. N. T. E. K. Zhiyuan Chen, "Exploration of the Effectiveness of Expectation Maximization Algorithm for Suspicious Transaction Detection in Anti-Money Laundering," *IEEE International Conference on Open Systems,* 2014.

[16] Q. Deng, "Detection of Fraudulent Financial Statements," *IEEE - Granular Computing 2009,* 2009.

[17] O. O. ,. O. O. Abiodun. Modupe Oludayo, "Exploring Support Vector Machines and Random Forests to Detect Advanced Fee Fraud Activities on Internet," *IEEE International Conference on Data Mining Workshops,* 2011.

[18] "CULS, Cornell University Law School, White-Collar Crime: an overview," [Online]. Available: http://topics.law.cornell.edu/wex/White-collar_crime (2009).

[19] W. ·. J. L. ·. L. C. ·. Y. O. J. Chen, "Effective detection of sophisticated online banking fraud on extremely imbalanced data," *Springer Science+Business Media,* 2012.

[20] "Cornell University Law School : Legal Information Institute," [Online]. Available: https://www.law.cornell.edu/wex/credit_card_fraud.

[21] L. A. V. D. A. M. d. C. F. S. M. Emanuel Mineda Carneiro, "Cluster Analysis and Artificial Neural Networks : A Case Study in Credit Card Fraud Detection," *International Conference on Information Technology - New Generations IEEE,* 2015.

[22] *Report of the National Commission on Fraudulent Financial Reporting,* New York: American Institute of Certified Public Accountants (AICPA), National Commission on Fraudulent Financial Reporting (Treadway Commission), 1987.

[23] V. R. ,. G. R. R. ,. I. B. P. Ravisankar, "Detection of financial statement fraud and feature selection using data mining techniques," *Decision Support Systems- Elsevier,* 2011.

[24] W. X. T. Xinyang Li, "How to Protect Investors? A GA-based DWD Approach for Financial Statement Fraud Detection," *IEEE International Conference on Systems, Man, and Cybernetics,* 2014.

[25] R. M. Manoj Kumar, "Unsupervised Outlier Detection Technique for Intrusion Detection in Cloud Computing," *International Conference for Convergence of Technology IEEE,* 2014.

[26] L. B. T. F. M. B. Mohammad Behdad, "On XCSR for electronic fraud detection," *Springer-Verlag ,* 2012.

[27] L. Z. a. S. X. Han Tao, "Insurance Fraud Identification Research Based on Fuzzy Support Vector Machine with Dual Membership," *International Conference on Information Management, Innovation Management and Industrial Engineering,* 2012.

[28] C. A. Melih Kirlidog, "A fraud detection approach with data mining in health insurance," *Procedia - Social and Behavioral Sciences,* 2012.

[29] J. Ariganello, What the Fraud? A Look at Telecommunications

Fraud and its Impact.

[30] K. T. J. V. T. Anna Leontjeva, "Fraud Detection: Methods of Analysis for Hypergraph Data," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* 2012.

[31] G. P. T. D. P. A. Theodoros Kapourniotis, "Scam and Fraud Detection in VoIP Networks: Analysis and Countermeasures Using User Profiling," *IEEE,* 2011.

[32] M. M. S. Hamid Farvaresh, "A datamining framework for detecting subscription fraud in telecommunication," *Engineering Applications of Artificial Intelligence.*

[33] S. P. Sharmila Subudhia, "Quarter-Sphere Support Vector Machine for Fraud Detection in Mobile Telecommunication Networks," *Procedia Computer Science,* 2015.

[34] G. A. Vipul Rawte, "Fraud Detection in Health Insurance using Data Mining Techniques," *International Conference on Communication, Information & Computing Technology (ICCICT),* 2015.

[35] J. P. M. A. E. G. F. V. Ll. Bermúdeza, "A Bayesian dichotomous model with asymmetric link for fraud in insurance," *Insurance: Mathematics and Economics Elsevier,* 2008.

[36] K. S. Y. S. K. T. S. K. A. a. M. M. Jawad Nagi, "Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines," *IEEE Transactions on Power Delivery Vol. 25,* 2010.

[37] A. M. M. K. S. Y. S. K. T. J. Nagi, "Non-Technical Loss Analysis for Detection of Electricity Theft using Support Vector Machines," *IEEE International Conference on Power and Energy,* 2008.

[38] C. S. Hilas, "An application of supervised and unsupervised learning approaches to telecommunications fraud detection," *Elsevier,* 2008.

[39] Z. Y. D. P. Z. Anisah H Nizar, "Detection Rules for Non Technical Losses," *IEEE,* 2008.

[40] J. O. P. P. A. M. A. C. P. José E. Cabral, "Fraud Detection System for High and Low Voltage Electricity Consumers Based on Data Mining," *Power & Energy Society General Meeting, 2009.,* 2009.

[41] C. S. Hilas, "An application of supervised and unsupervised learning approaches to telecommunications fraud detection," *Elsevier,* 2008.

[42] C. S. Hilas, "Designing an expert system for fraud detection in private telecommunications networks," *Expert Systems with Applications Elsevier,* 2009.

[43] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," *Elsevier,* 2014.

[44] H. T. a. A. N. R. Huang, "A Novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection," *International Conference on Computational Science, IEEE,* 2010.

[45] C. W. ·. D. J. H. ·. P. J. ·. D. W. ·. N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Springer Science+Business Media,* 2008.

[46] M. I. E. D. T. C. M. Hamdi Ozelik, "Improving a credit card fraud detection system using genetic algorithm," *International Conference on Networking and Information Technology,* 2010.

[47] A. N. Tatsuya Minegishi, "Detection of Fraud Use of Credit Card by Extended VFDT," *IEEE,* 2011.

[48] L. Lei, "Card Fraud Detection by Inductive Learning and Evolutionary Algorithm," *International Conference on Genetic and Evolutionary Computing IEEE,* 2012.

[49] R. N. Sherly K.K, "BOAT Adaptive Credit Card Fraud Detection System," *IEEE,* 2010.

[50] R. D. Mukesh Kumar Mishra, "A Comparative Study of Chebyshev Functional Link Artificial Neural Network, Multi-Layer Perceptron and Decision Tree for Credit Card Fraud Detection," *13th International Conference on Information Technology IEEE,* 2014.

[51] M. R. H. Leila Seyedhossein, "Mining Information from Credit Card Time Series for Timelier Fraud Detection," *5th International Symposium on Telecommunications,* 2010.

[52] G. S. A. S. L. V. Addisson Salazar, "Automatic Credit Card Fraud Detection based on Non-linear Signal Processing," *IEEE,* 2012.

[53] N. W. Wen-Fang YU, "Research on Credit Card Fraud Detection Model Based on Distance Sum," *International Joint Conference on Artificial Intelligence,* 2009.

[54] N. W. Chun-Hua JU, "Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum," *First International Workshop on Database Technology and Applications,* 2009.

[55] P. S. Masoumeh Zareapoor, "Applications of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," *Procedia Journal,* 2015.

[56] Y. Z. ,. X. Timothy Matti, "Financial Fraud Detection Using Social Media Crowdsourcing," *IEEE,* 2014.

[57] M. R. H. Soheil Jamshidi, "An Efficient Data Enrichment Scheme for Fraud Detection Using Social Network Analysis," *IEEE,* 2012.

[58] G. M. Qingshan Deng, "Combining Self-Organizing Map and K-Means Clustering for Detecting Fraudulent Financial Statements," *IEEE - Granular Computing 2009,* 2009.