International Journal of Engineering & Technology, 7 (4) (2018) 5862-5868

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.15696
Research paper

A survey on machine learning techniques for fraud detection in healthcare

Shamitha S. K [1] *, V. Ilango [2]

[1] Research Scholar, New Horizon College of Engineering, Bangalore
[2] Professor, CMR Institute of Technology, Bangalore
**Corresponding author E-mail: shamithashibu@gmail.com*

Abstract

An exponential upward change in fraud occurrence has resulted in billions of dollars loss in the world economy. Newer techniques in fraud detection in healthcare domain are continuously evolving and are put into practice in many business fields. In healthcare Fraud detection, user behavior is monitored to analyze and find any suspicious or undesirable behavior and to avoid the same. Undesirable behavior could be anything like crime, fraud, unwarranted intrusion or any other kind of default. The goal of this paper is to provide a comprehensive review of different types of fraud and fraud detection techniques used in last two decades.

Keywords: *Fraud; Fraud Detection Techniques; Healthcare; Machine Learning Techniques.*

1. Introduction

Patient medical records, statutory documents, and record keepings has helped the healthcare industry accumulate large amount of data which is mostly kept as hard copy. Recently there has been urge towards digitization of these records in rapid pace. Fraud in healthcare industry is defined as knowingly making a fictitious claim, inflating a claim or adding extra items to a claim, or being in any way dishonest with the intention of gaining more than legitimate entitlement. Healthcare analytics depicts the social insurance exercises that can be attempted because of information gathered from four distinct zones inside medicinal services; claims and cost information, pharmaceutical and innovative work (Research and development) information, clinical information (gathered from electronic therapeutic records (EHRs)), and patient role behavior and sentiment data (patient behaviors and inclinations), retail buys for instance information those are caught in running stores.

Till recently, healthcare services had been utilizing the preservationist strategy for determination and treatment, where most specialists relied upon their individual experience and aptitudes in diagnosing maladies in patients, bringing about a less exact and patient-driven. Digitization, rising rates of perpetual ailments, expanded populace, progression in innovation, the requirement for evidence-based medication, powerlessness to outgrowth and get knowledge from consistently expanding heterogeneous medicinal information are a portion of the drivers for embracing health care analytics. Insurance plays a central role in the health care field. More than 80 percentage of healthcare expenditures are funded by insurance companies, either public or private. Insurance thus offers the money that motivates and cares the health care system. The inspiration for this paper comes from a real-life fraud in health care.

Various statistical survey on losses incurred by fraud in different countries has been referred on several literatures and articles. In developed country like USA, around $47.9 billion of the country's expenditure was lost due to fraud. It was estimated that the false claims itself was almost 15% of the total claims[1], and developing country like India, its private expenditure amounts are around 4% of the GDP[2], out of this health insurance accounts for about 5-10% of the total expenditure. It is found that there are not much literatures for fraud detection based on Indian data. While coming into public health scheme during last few decades, it has increased to almost 55 million people. Nearly two thirds of its population are those in Below Poverty Line (BPL). Considering all these concerns India is losing approximately 600 to 800 crores of rupees incurred on fraudulent claims annually[3]. The Association of British Insurers claims that "fraudulent claims in healthcare industry in UK insurance costs over 1 billion a year and fraudsters continuously develop new types of scams" [4]. In Turkey, losses are accumulated by the insurers as higher premiums. Estimated fraud on insurance in Canadian insurance industry is around 1.3 billion Canadian Dollars every year which transforms to about 10-15% of the claims paid out in Canada (Gill, 2009).

Electronic claims processing has been increasingly actualized nowadays which performs reviews and surveys of cases information naturally. These electronic frameworks recognizes zones which requires extraordinary consideration, for example, incorrect or deficient information input, copy claims, and medicinally no secured administrations. In spite of the fact that these frameworks might be utilized to recognize certain sorts of extortion, since these framework does not have any intense worked in misrepresentation discovery procedures their capacities in distinguishing extortion are typically restricted since the identification fundamentally depends on pre-characterized basic standards specified by space specialists. To accomplish a powerful extortion discovery, it has turned into a need to grow increasingly complex antifraud approaches joining information mining, machine learning or different techniques. New proposed approaches center around more muddled assignments, for example, programmed taking in of extortion designs from information, determine "misrepresentation probability" of each case to organize some suspicious cases, and

Copyright © 2018 Shamitha S. K, V. Ilango. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

distinguish new sort of extortion which were not already reported [4-5].

The perception of increase in fraud during last few years has made people to invest more on fraud detection schemes. The report by "State Insurance Fraud" suggests that 50 percent of data has increased slightly on last three years. Around 76 percent of insurers says that detecting claims fraud is the primary aim of anti-fraud technology in healthcare because claims fraud is increasing dramatically compared to other frauds in healthcare. That is up from 71 percent in 2014 and 65 percent in 2012. The report also suggests that from Anti-fraud technology which is currently employed, Predictive modeling has increased from 30% to 55% in last three years. There is only a slight variation in use of anomaly detection. Technologies based on automated red flags/business rules had reached its peak that means from 55% to around 80%[6][7].

The paper is organized into four sections. Section two gives an overview of fraud. Section three explains various types of fraud in healthcare. Section four reviews various types of fraud detection techniques focusing on health care domain. And section five concludes the paper.

2. Fraud

Fraud is defined as "criminal deception" as defined by Concise Oxford Dictionary; which means to use `false representation to gain an unfair advantage [8]. Fraud, in earlier days, was restricted only to money laundering services which was easier to bring about. Nowadays, due to digitalization in almost all sectors has made an uncontrollable increase in fraudulent activities, making fraud detection more important than ever [8]. Fraud can be internal or external. Internal fraud is normally committed by employees inside the organization, mostly by means of theft of cash or stock, providing services for near ones without incurring any cash, allowing companies to use bad credit, supplying receipts for refunds etc. Insurance is a contract in which a person receives financial protection or reimbursement against losses from an organization. Fraud in insurance may occur at any stage like during buying, selling or while staking a claim[9 - 11]. On a survey conducted by "The State of Insurance Fraud Technology", more than half of the people stated that the fraud has increased. Only 2 percentage states that it has decreased[6]. Healthcare frauds and its detection using machine learning techniques will be discussed further in the paper.

3. Types of fraud in health care

Healthcare fraud involves filing dishonest health claims for profit. Fraud involves an intentional effort to misrepresent facts to make a profit whether financial or material [12]. Among the articles reviewed fraud is broadly classified into two, single fraud and multiple fraud. Kickbacks are the most common type of fraud, which is a form of bribery [13]. Pharmacist fills the prescription for those brand in which he can yield a bonus from the particular company[14-16]. Sheehan and Goldner in their article has described about Self-Referral Fraud, which is another type of kickback Fraud it involves referring the patients to the clinic if a bribe is being paid for that[17]. A patient may try to obtain prescription multiple times by visiting multiple doctors for the same disease, this type of fraud is called doctor shopping[18]. Identity fraud is another type of fraud where in uninsured individual assumes the identity of an insured person to obtain services[19]. One of the most prevalent fraud activity possessing an expensive service when the actual are only mediocre is upcoding. Phantom billing is said to be a kind of upcoding fraud in which billing is done for the services not provided. There are certain cases where claims are submitted for a service provided based on already stated diagnosis, such type of fraud are done to falsely prescribe medicines [16] [20]. Price manipulation by pharmaceutical companies is another way of fraudulent activity. If the provider is aware that the patient here is holding an insurance they may move across the street where which the charge would be higher. Insurance amount is claimed based on the stated diagnosis, the diagnosis can also be manipulated by providing false diagnosis and falsely prescribe certain medicines to patients, such kind of fraud is discussed by (Ogunbanjo) on his paper. Some hospitals make the patient to be hospitalized for more days and providing unnecessary care though the patient will be in a condition to be discharged [21], which describes how the healthcare service providers maximize their services and there by billing for those services which are of least necessity. Fig 1 illustrates various fraud types in health insurance as referred in multiple articles.

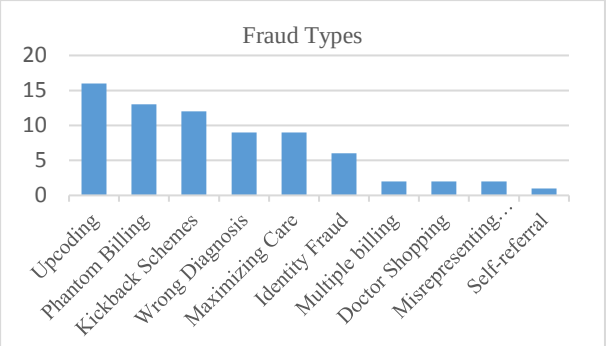


Fig. 1: Occurrence of Health Insurance Fraud Types in Literature.

4. Fraud detection techniques

With the growth in modern technologies fraud has been increased drastically, which may lead to a sluggish economy. According to a review conducted by "National fraud Authority", techniques used to commit frauds can be divided into three major areas such as victim selection techniques, perpetration strategies, and detection[22]. Fraud detection methods are continuously upgraded or new methods are being invented to defend criminals in pursuing their innovations. Machine learning methods are proved as most convenient methods for detecting fraud[23]. It also tries to learn pattern from the data and also provides specification of fraud possibility for each and every cases so that any suspicious cases can be easily speculated. Examples of supervised statistical methods are neural networks, decision trees, Bayesian networks, and genetic algorithms[24]. Neural networks are considered to be best method for data structure with non-linear relationship. To find fraud that deviates from the group unsupervised statistical method can be used. Examples of unsupervised methods are Electronic Fraud detection (EFD), Smart sifter[25]. Machine learning methods include Data mining techniques and statistical techniques. A comparative study conducted by [26] wipro states that machine learning techniques procure improved predictive accuracy and also enables higher coverage with low false positive rates. Data mining techniques are most widely used techniques for detecting fraud in health insurance when the data is in huge quantity.

4.1. Data mining techniques

Data is stored in real world databases and as time moves this amount continues to grow faster. Table III gives a detailed review on different fraud types and fraud detection techniques discussed in various literatures over past two decades.

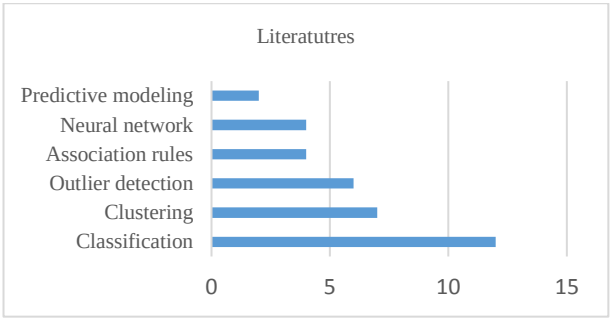


Fig. 2: Occurrences of Fraud Detection Methods for Healthcare from Various Literature.

knowledge in such database which automatically filters through immense amounts of data to find hidden patterns, bring out valuable perceptions and make predictions. Data mining techniques have a tendency to learn models from data. Learning approaches on data mining models are divided into two types supervised and unsupervised learning

4.1.1. Supervised learning

Supervised data mining techniques are used when training data is available. In case of fraud detection, a particular kind of fraud is identified based on the previously known or identified fraud. In health Insurance fraud detection, the pattern or label may be "legitimate", "fraudulent" claims. In case of a new claim it can be compared with the defined label to find any occurrence of a suspicious match[3].Supervised technique is a best-defined technique for pattern classification

4.1.2. Unsupervised learning

Unsupervised data mining technique does not have any predefined labels or training data set. While comparing to supervised learning, unsupervised learning is not restricted to patterns or labels, so when used in fraud detection it can easily identify old and new kind of frauds. Clustering or outlier detection method can be used since there is no training data available. Table I describes various data mining techniques with the methods and algorithms. Learning classifiers such as Boosting, Modified Randomized Under sampling (MRU), and Adjusted Minority Oversampling (MO) can also be used for prediction. Self-Organized Maps, Multilayer perceptron genetic algorithm combined to KNN can also be used as an AI or predictive modeling technique to detect wrong diagnosis, a kind of fraud type[27]. Social Network Analysis, a method of predictive modelling calculates propensity scores and can forecast automatically that which data will be fraudulent. The Incidence of different healthcare fraud detection methods from various literatures
are shown in the Fig. 2. Among the literatures reviewed majority of the literatures uses datamining techniques such as Clustering, Outlier detection, classification, Association rules etc.

Table 1: Data Mining Techniques Its Methods and Algorithms

Techniques	Methods	Algorithms
Supervised Data Mining		Support vector machine, Gaussian, Linear kernels, Naïve Bayes, decision tree, Multiple Criteria Linear Programming Polymorphous(M-of-N) logic, Self-organizing map Association rule
	Classification	Neural networks, decision trees, fuzzy logic, genetic algorithms
	Clustering	SAS-EM, CLUTO, k-means Algorithm Nonnegative matrix factorization
Un Supervised Data Mining		Regression analysis, Distance analysis Decision trees, Distance analysis, Density estimation Pairwise comparison
	Outlier detection	Electronic Fraud detection (EFD), Smart sifter

5. Conclusion

This paper provides a systematic literature review of various fraud types in healthcare and its detection techniques. The whole review is put across in a tabular format. Much of the work on this domain is done in recent years and yet more work remains. When dealing with fraud and fraud detection, understanding your enemy and find the level of loss and victory is most important. Healthcare fraud is a growing type of crime. The ultimate aim should be in developing a high-quality and cost-effective fraud detection system to deal with all kind of healthcare fraud. Efforts are made to classify types of fraud and various fraud detection techniques. Most of the detection techniques prevailing till recent years were surrounding data mining, but recently there has been a rising significance on Artificial Intelligence techniques as well. This is due to its power of deep learning or predictive modeling. AI is useful in solving complex types of fraud with a higher level of abstraction. Future work will be focusing on Artificial intelligence techniques for fraud detection in healthcare

Table 2: Literature Survey on Fraud Detection Techniques Based on Fraud Types

Author	Type of fraud	Techniques	Methods	Description
Rawte et al. [3]	Identity fraud	Clustering	Hybrid Evolving Clustering Method and SVM	The overall approach worked as follows, the insurance claims will be clustered using ECM according to the disease type and then classified using Support Vector Machine to detect for any duplicate claims. The author claims that the hybrid approach of ECM clustering and SVM was able to include new unknown frauds when compared to traditional k means clustering.
Laleh et al. [5]	Insurance Fraud	Data Mining Techniques	Batch time techniques and Real time techniques.	The author explains how supervised and semi supervised methods can be used to detect fraud. Supervised methods such as Neural network and Bayesian, uses STAGE Algorithm and Back propagation algorithm. A data preprocessing is followed by that which uses rough set algorithm. After that Multi algorithmic approach and adaptive CBR is used for fraud filtering and classification.
Sithic et al. [11]	Identity fraud, Falsifying benefits	Data Mining Techniques	Clustering, Classification, Regression, Visualization.	The paper categorizes data mining techniques and uses in fraud detection. It also explains various types of fraudsters in insurance sector. The paper states that India is losing around 600 crore from its economy due to the fraud.
Li et al. [24]	Healthcare fraud	Neural Networks, Classification, Outlier detection	Bayesian Networks, Genetic Algorithms, Smart Sifter, Electronic Fraud Detection.	Supervised and Unsupervised methods of fraud detection is explained. Neural Networks, Bayesian Networks and Genetic Algorithms are supervised methods and Smart Sifter and Electronic Fraud Detection are unsupervised methods
Ana-Ramona et al. [28]	Identity Fraud	Predictive modeling	Social Network Analysis,	Predictive Modelling technique is used for fraud detection by Social Network Analysis method, Predictive models uses probability that data can be found by calculating fraud propensity scores and can automatically shows which data will be fraudulent. In SNA suspi-

Anbarasi et al. [29]	Upcoding	Outlier detection method	Pairwise comparison method	cious components are detected on the basis of shared characteristics and there will be defined set of indicators. The model combines proactive and retrospective analysis to reduce time consumption than other models. It uses the pairwise comparison method to approve and reject claims in an acceptable time by showing the risk degree of an actor. Finally outlier detection methods is used to filter the irrelevant data by using metrics or predictors which deviate from the usual patterns. The true positive rates of the model varies between 89.5% to 94.5% and true negative rates between 52.8% and 83.4%.
Bauder et al. [30]	Phantom billing	Classification	Naïve Bayes algorithm	A multinomial Naïve Bayes classifier evaluated using 5 fold cross-validation and three performance matrices such as recall, precision and F-score is built. The classes which are well predicted will only be used for anomaly detection, Proper conditions where determined for finding the physicians fraud. F-score values were found and categorized into three groups, higher F-score indicated good results, lower F-scores indicates the chances of overlapping the procedure
Borse et al. [31]	Phantom billing	Classification	Naïve Bayes	Four attributes are considered for the test in the paper, the test data with the values are passed to check whether the entry is fraudulent or not the inference is that service providers had manipulated the actual count of services that were to be provided to the patients and if the output comes as not fraud, then the inference would be that the test data was not fraudulent and no service provider's fraud is detected. Naïve Bayes was found a better solution for detecting Service fraud.
Brownell et al. [32]	Fraudulent Claims	Prediction	Ethnographic approach	Ethnographic approach uses interviewing, participant observation, non-participant observation, and practical hands-on experience, to build up a depiction of procedure of fraud detection as it occurs in the complex environment of insurance practice. Using an ethnographic approach, it is possible to elicit from investigators detailed expertise on how best to support fraud detection by less experienced claims handlers at front end of the claims process.
Chandola et al. [33]	Anomaly detection	Classification, Artificial Intelligence	Classification based Anomaly detection, Multilayered perceptron, Auto-Associative Networks	The paper provides a comprehensive survey on Anomaly detection, Classification based, Nearest Neighbor, and Cluster based techniques, Non parametric techniques and Statistical Anomaly detection techniques along with its computational complexities and advantages are discussed in the paper. For each type of anomaly detection a unique notion of normal and anomalous data are identified.
Chen et al. [34]	Physician Collusion	Community detection Algorithm	Spectral Analysis	The paper proposes a novel fraud detection technique which works using community partition algorithm through spectral analysis for detecting external frauds with an unsupervised learning approach. A Gap-Cut algorithm based on Spectral Analysis network can be divided into an unknown number of communities. Gap-Cut algorithms are compared with other community finding alg. which gave comparable good results and had greater flexibilities.
CLIFTON PHUA et al. [35]	Healthcare fraud	Supervised Algorithms	True positive rates	The author explains following algorithms for fraud detection, supervised algorithm uses either "true positive" rates or "accuracy at a chosen threshold", semi Supervised algorithm such as Anomaly detection, conditional entropy. Relative conditional entropy etc. Supervised algorithms are mainly used to process labelled data, Author also explains disadvantages of using labelled data such as if training data is incorrect it may result to bias, to overcome this it advised to combine training data and processed by multiple unsupervised algorithms.
Coderre [36]	Kickback fraud	Digital Analysis	Benfords Law	The paper uses benfords law, which when compared against Jonathan's kickback scheme, which had a ratio in between 0.7 to 1.30.
Cui et al. [37]	Wrong Diagnosis	Data Mining Approaches	GM-FP, HITS algorithm.	A combined approach on graph mining with frequent pattern (GM-FP) is introduced. It analyses every record to identify CPB(Copying prescription Behaviour).A network graph was used to depict the relationship between doctor and patient along with its weight, this process will be done iteratively using HITS algorithm. In addition to this the paper has also calculated prescription copying behavior in the treatment sequence, It was also found that GM-FP outperformed all the existing algorithm in terms of F-score.
Derrig [38]	Phantom billing	Classification and clustering, Fuzzy logic	Claim sorting algorithm, PRIDIT, Tune investigation	The procedure works as follows, claims are sorted based on their scores, and moved to their respective bins. In case of multidimensional scores, bin assignments can be made by fuzzy clustering. PRIDIT score will be used to show the gain or loss of changing data patterns. To optimize the detection and deference of fraud and abuse, tune investigative approach is used.
Francis et al. [39]	Phantom billing	Machine Learning	Linear Support Vector Machine	The research was focusing on medical fraud, filling of dishonest claims and billing patterns that occur in US. Area of work is on workers compensation bills, execution of the bills was performed by running multiple instances of each task in parallel. This method yielded best efficiency and also ease on handling load. Use of SVM resulted a better accuracy on training and classification run time.
Gangopadhyay et al. [40]	Self-Referral	Clustering	Community Detection Algorithm	The paper focuses on connection based similarities, which will help to find the connection between the parties. These connections are reflected by building a matrix. Using this matrix the communities which are most suspicious could be found. For large data bases,

				results are calculates and aggregated using the segments. The algorithm was able to execute around 50000 physician in 1 minute.
H. Peng et al. [41]	Unbundling/ Upcoding	Neural Networks and Clustering	Pharmacopoeia spectrum tree	A three layer of Multi-Layer Perceptron feed forward method is introduced, which is based on neural network and clustering technique. To get a reasonable fraud factor the author builds a pharmacopoeia spectrum tree and use the neural network using NN and clustering technique for over-fitting and under-fitting algorithm. As per author it proved to be far improved version of other unsupervised clustering methods.
Jiwon Seo et al. [42]	Improper coding	Classification	Page Rank algorithm	The algorithm identifies medical providers whose prescribed medical procedures are significantly different from those of other providers. A similarity graph is generated with nodes as providers and edges as edges indicate whether two providers are similar. The algorithm was executed with different threshold values ranging from 0.65 to 0.8, the results appeared similar. The detection algorithm was applied to different specialties which gave clear and accurate.
Kirlidog et al. [43]	Medical fraud	Classification	Support Vector Machine	Probability or likelihood of each record is calculated and if probability is higher than 50% than it will be marked anomalous. The main task will be aimed to narrow the target for detecting claims. The technique identified 6595 records with probabilities ranging from 50.0% to 67.3 %. The
Konijn et al. [44]	Multiple Billing, Unbundling	Outlier detection Approach	LOf scores, z scores	Research was focusing on three types of deviations such as unusual prescriptions, typo errors, and unusual number of expensive patterns. Modified LOF score is used to calculate fraudulent pattern in case of multiple billing. The results showed that the number of pills were much lower than that of claimed, which states the unbundling fraud. To identify the mistakes made by pharmacies z scores are used.
Lata et al. [45]	All fraud types	Data Mining Techniques	AINDS, Bayesian, Smart Sifter, Electronic Fraud detection, Neural Network	Impact of data mining techniques, credit card fraud and Intrusion detection techniques on health insurance fraud is discussed in the paper. Fraud detection techniques in telecommunications and healthcare are discussed. For healthcare fraud many techniques are discussed among them neural network and decision tree gave better results
Lin et al. [46]	Up coding	Classification	decision tress, rule based	The research worked as follows, A decision tree is formed first keeping the fraud as target variable followed by data transformation, to classify and structuralize group, relevant information is retrieved based on the information from the decision tree, actions are taken based on the claims. This model was compared with the older model and the findings suggest that the dependence of high input labor cost is being relieved.
Liu et al. [47]	Falsifying benefits/ claims	Clustering	Geo-Location clustering model	The proposed model divides the data set into three clusters, the model detects claims with extreme payment amount as well as with large payment amount, by calculating maximum, minimum, median distance and standard deviation between the points in the cluster. The author claims the model is best suitable for preliminary Analytic procedure.
Lu et al. [48]	All health care Fraud	Outlier detection method	Adaptive Benford Algorithm.	The proposed algorithm works with incomplete data as well, if any large anomalies are detected it will be reported as fraud. The algorithm when compared to basic benford's law produces a more precise set of anomalies especially for forensic auditors to analyze fraud.
Lu et al. [49]	All health care fraud	Outlier detection	Adaptive Benfords Law	Adaptive Benfords law was an extension to outlier detection methods. Data sets were analyzed and compared with three different benford Law digit frequency distributions. The method was enhanced with a reinforcement learning model in order to link together anomalous outliers to build case of fraud.
Musal [50]	Unbundling, Upcoding	Clustering and Regression	Peer comparison, Distance Analysis	Clustering procedures as well as regression for geographical analysis of possible Medicare fraud. The proposed model uses clustering procedure to group areas. Regression analysis is used to achieve this discrimination. By this method almost all healthcare provider will be flagged and can possibly identify the fraud easily.
Sadiq et al. [51]	Up coding	Classification	Patient Rule Induction Methods	PRIM generates the decision rules to delineate a region by recursively peeling non-bump regions from the result and thus the remaining subsets are regarded as bumps. PRIM was evaluated against several algorithms and achieved better F-score and accuracy results because it characterizes the input space and then classifies new instances.
Steventon et al. [52]	Data Entry Errors	Tele monitoring Technique	Tele-F	The paper is a case study on Tele-F, which is a tele monitoring to improve Heart failure outcomes. The data monitored is the weight of patients which is manually entered by the officials, which are tend to entered wrongly by rounding up the digits or misrepresentation.114, 867 weight readings were reported out of which 18.6% were affected by end-digit preference.105 patients demonstrated end-digit preference on 14.9% of those who submitted data. Patients with end-digit preference generated an average 2.9 alerts to the Tele-F system over the six-month trial period.
Tagaris et al. [53]	Phantom Billing	Classification	Rule based classification	The paper indicates predefined set of rules entered in the system reporting all suspicious cases, the data is checked along with electronic data entry of prescription written by doctors. The application

Thornton et al. [54]	Up coding	Outlier detection method	Data driven approach	was fed 20,000 prescriptions, various statistical forms of fraudulent behavior were examined. "Insured person' Total number of diagnoses" indicated that 2 percentage of the insured persons had 17 diagnosis same year which indicates a chances of fraud. The paper provides a detailed description with facts on the method for applying outlier detection for health insurance fraud. The outlier model was evaluated with actual data, the fraud found was 17% as compared to prior successive rate of 10%.
Viaene et al. [55]	Claims fraud	Classification	AdaBoost naïve bayes, Adaptive Resample and combined algorithm	AdaBoost naïve bayes combines the advantages of boosting, flexibility and representational attractiveness of the probabilistic weight of evidence scoring framework can be applied for the diagnosis of insurance claim fraud. AdaBoosted weight of evidence scoring frame- work offers readily accessible and naturally interpretable decision support and allows for flexible human expert interaction and tuning
Y. Peng et al. [56]	Healthcare fraud	Clustering	SAS EM and CLUTO	Clustering methods like SAS EM and CLUTO to detect health fraud is discussed in the paper. Experiments conducted in the paper describes that CLUTO takes less computation time than SAS EM, and in CLUTO records were concentrated in one cluster and in SAS EM it was distributed.
Lee et al. [57]	Skewed data distributions	Neural Networks, Classification	Back Propagation, Naïve Bayesian and C4.5 algorithms	The proposed method compares meta-learning approach against C4.5 and SMOTEing without partition, which is a sampling approach. Partitioning algorithm achieves marginally higher cost savings by the given threshold.
Pejic-Bach [58]	Wrong Diagnosis	Neural Networks, Fuzzy sets,	Self-Organizing Maps and KNN	The paper explains fraud types and detection methods in various domains including healthcare. Author explains IS methods for detecting wrong diagnosis such as SOMs, Multilayer perceptron, genetic algorithm combined to KNN. Several literatures on fraud detection based on these methods are also listed in the paper.

References

[1] M. P. Pawar, "Review on Data Mining Techniques for Fraud Detection in Health Insurance," vol. 3, no. 2, pp. 1128–1131, 2016.

[2] McKinsey-CII, "India Health care: Inspiring possibilities, challenging journey," no. December, p. 34, 2012.

[3] V. Rawte and G. Anuradha, "Fraud detection in health insurance using data mining techniques," in 2015 International Conference on Communication, Information & Computing Technology (ICCICT), 2015, pp. 1–5. https://doi.org/10.1109/ICCICT.2015.7045689.

[4] N. J. Morley, L. J. Ball, and T. C. Ormerod, "How the detection of insurance fraud succeeds and fails," Psychol. Crime Law, vol. 12, no. 2, pp. 163–180, 2006. https://doi.org/10.1080/10683160512331316325.

[5] N. Laleh and M. Abdollahi Azgomi, "A taxonomy of frauds and fraud detection techniques," Commun. Comput. Inf. Sci., vol. 31, pp. 256–267, 2009. https://doi.org/10.1007/978-3-642-00405-6_28.

[6] The State of Insurance Fraud Technology," no. November 2016.

[7] K. K. Tripathi and M. A. Pavaskar, "Survey on Credit Card Fraud Detection Methods," Int. J. Emerg. Technol. Adv. Eng., vol. 2, no. 11, p. 721, 2012.

[8] T. P. Bhatla, V. Prabhu, and A. Dua, "Understanding Credit Card Frauds," Cards Bus. Rev., vol. 1, no. 6, pp. 1–15, 2003.

[9] A. Georgia, "Telecommunications fraud," Ind. Eng. IE, vol. Jun2007, V, p. 2p; 2 Color Photographs, 2007.

[10] D. Olszewski, "A probabilistic approach to fraud detection in telecommunications," Knowledge-Based Syst., vol. 26, pp. 246–258, 2012. https://doi.org/10.1016/j.knosys.2011.08.018.

[11] H. Sithic and T. Balasubramanian, "Survey of Insurance Fraud Detection Using Data Mining Techniques," Int. J. Innov. Technol. Explor. Eng., vol. 2, no. 3, pp. 62–65, 2013.

[12] S. S. Waghade, "A Comprehensive Study of Healthcare Fraud Detection based on Machine Learning," vol. 13, no. 6, pp. 4175–4178, 2018.

[13] S. Pandit, S. Wang, and C. Faloutsos, "NetProbe : A Fast and Scalable System for Fraud Detection in Online Auction Networks NetProbe : A Fast and Scalable System for Fraud Detection in Online Auction Networks," Proc. 16th Int. Conf. World Wide Web, pp. 201–210, 2007. https://doi.org/10.1145/1242572.1242600.

[14] J. K. Taitsman, "Educating Physicians to Prevent Fraud, Waste, and Abuse," N. Engl. J. Med., vol. 364, no. 2, pp. 102–103, 2011. https://doi.org/10.1056/NEJMp1012609.

[15] J. Sheehan and J. Goldner, "Beyond the Anti-Kickback Statute: New Entities, New Theories in Healthcare Fraud Prosecutions," J. Health Law, vol. 242, no. c, pp. 1419–1420, 2007.

[16] CMS, "DEPARTMENT OF HEALTH AND HUMAN SERVICES Centers for Medicare & Medicaid Services," no. October 2016.

[17] A. Rashidian, H. Joudaki, and T. Vian, "No evidence of the effect of the interventions to combat health care fraud and abuse: A systematic review of literature," PLoS ONE. 2012. https://doi.org/10.1371/journal.pone.0041988.

[18] J. Carlson, "Painful side effects," Mod. Healthc., 2013.

[19] J. F. Dube, "Fraud in Health Care and Organized Crime," Med. Health, vol. 94, no. 9, pp. 268–269, 2009.

[20] G. A. Ogunbanjo and D. K. van Bogaert, "Ethics in health care: healthcare fraud," South African Fam. Pract., vol. 56, no. 1, pp. 10–13, 2014.

[21] L. Morris, "Perspective: Combating fraud in health care: An essential component of any cost containment strategy," Health Aff., vol. 28, no. 5, pp. 1351–1356, 2009. https://doi.org/10.1377/hlthaff.28.5.1351.

[22] C. L. and J. T. Mark Button, "Fraud typologies and victims of fraud," pp. 1–40, 2010.

[23] Y.-P. Huang, "Survey of Fraud Detection Techniques," 2004, no. September, pp. 749–754.

[24] J. Li, K.-Y. Huang, J. Jin, and J. Shi, "A survey on statistical methods for health care fraud detection," Health Care Manag. Sci., vol. 11, no. 3, pp. 275–287, 2008. https://doi.org/10.1007/s10729-007-9045-4.

[25] K. Yamanishi and G. Williams, "On-line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms," pp. 320–324, 2000. https://doi.org/10.1145/347090.347160.

[26] WIPRO, "Comparative Analysis of Machine Learning Techniques for Detecting Insurance Claims Fraud," WIPRO Ltd., 2015.

[27] A. M. Mubarek and E. Adali, "Multilayer perceptron neural network technique for fraud detection," 2017 Int. Conf. Comput. Sci. Eng., pp. 383–387, 2017. https://doi.org/10.1109/UBMK.2017.8093417.

[28] A. F. Ana-Ramona BOLOGA, Razvan BOLOGA, "Big Data and Specific Analysis Methods for Insurance Fraud Detection," Database Syst. J., vol. 4, no. 4, pp. 30–39, 2013.

[29] M. S. Anbarasi, "Fraud detection using outlier predictor in health insurance data," no. Icices, 2017. https://doi.org/10.1109/ICICES.2017.8070750.

[30] R. A. Bauder, T. M. Khoshgoftaar, A. Richter, and M. Herland, "Predicting Medical Provider Specialties to Detect Anomalous Insurance Claims," 2016. https://doi.org/10.1109/ICTAI.2016.0123.

[31] N. Borse and N. Maitre, "HEALTH CARE INSURANCE FRAUD DETECTION: A DATA MINING PERSPECTIVE," no. 2, pp. 52–56, 2015.

[32] H. Brownell, R. Griffin, E. Winner, O. Friedman, and F. Happé, "Address for correspondence," pp. 1–32, 2000.

[33] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv., vol. 41, no. September, pp. 1–58, 2009. https://doi.org/10.1145/1541880.1541882.

[34] S. Chen and A. Gangopadhyay, "A novel approach to uncover health care frauds through spectral analysis," Proc. - 2013 IEEE Int. Conf. Healthc. Informatics, ICHI 2013, pp. 499–504, 2013. https://doi.org/10.1109/ICHI.2013.77.

[35] K. S. & R. G. CLIFTON PHUA*, VINCENT LEE, "A comprehensive survey of data mining-based accounting-fraud detection research," 2010 Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2010, vol. 1, pp. 50–53, 2010.

[36] D. Coderre, "Fraud Detection Using Digital Analysis," EDP Audit. Control. Secur. Newsletter (EDPACS), vol. 27, no. 3, pp. 1–8, 1999. https://doi.org/10.1201/1079/43249.27.3.19990901/30268.1.

[37] H. Cui, Q. Li, H. Li, and Z. Yan, "Healthcare Fraud Detection Based on Trustworthiness of Doctors," 2016. https://doi.org/10.1109/TrustCom.2016.0048.

[38] R. a Derrig, "Insurance fraud," vol. 69, no. 3, pp. 271–287, 2002. https://doi.org/10.1111/1539-6975.00026.

[39] C. Francis, N. Pepper, and H. Strong, "Using support vector machines to detect medical fraud and abuse," 2011 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 8291–8294, 2011. https://doi.org/10.1109/IEMBS.2011.6092044.

[40] A. Gangopadhyay and S. Chen, "Health Care Fraud Detection with Community Detection Algorithms," 2016 IEEE Int. Conf. Smart Comput., pp. 1–5, 2016. https://doi.org/10.1109/SMARTCOMP.2016.7501694.

[41] H. Peng and M. You, "The Health Care Fraud Detection Using the Pharmacopoeia Spectrum Tree and Neural Network," pp. 2008–2013, 2016. https://doi.org/10.1109/TrustCom.2016.0306.

[42] J. Jiwon Seo and O. Mendelevitch, "Identifying frauds and anomalies in Medicare-B dataset.," Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., vol. 2017, pp. 3664–3667, 2017.

[43] M. Kirlidog and C. Asuk, "A Fraud Detection Approach with Data Mining in Health Insurance," Procedia - Soc. Behav. Sci., vol. 62, pp. 989–994, 2012. https://doi.org/10.1016/j.sbspro.2012.09.168.

[44] R. M. Konijn and W. Kowalczyk, "Finding fraud in health insurance data with two-layer outlier detection approach," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6862 LNCS, pp. 394–405, 2011. https://doi.org/10.1007/978-3-642-23544-3_30.

[45] L. N. Lata, I. A. Koushika, and S. S. Hasan, "A Comprehensive Survey of Fraud Detection Techniques," Int. J. Appl. Inf. Syst., vol. 10, no. 2, pp. 26–32, 2015. https://doi.org/10.5120/ijais2015451471.

[46] K.-C. Lin and C.-L. Yeh, "Use of Data Mining Techniques to Detect Medical Fraud in Health Insurance," Int. J. Eng. Technol. Innov., vol. 2, no. 2, pp. 126–137, 2012.

[47] Q. Liu and M. Vasarhelyi, "Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information," 29th WORLD Contin. Audit. Report. Symp., 2013.

[48] F. Lu and J. E. Boritz, "Detecting Fraud in Health Insurance Data: Learning to Model Incomplete Benford's Law Distributions," pp. 633–640, 2005. https://doi.org/10.1007/11564096_63.

[49] F. Lu, J. E. Boritz, and D. Covvey, "Adaptive Fraud Detection using Benford's Law," Adv. Artif. Intell. Proc., vol. 4013, pp. 347–358, 2006. https://doi.org/10.1007/11766247_30.

[50] R. M. Musal, "Two models to investigate medicare fraud within unsupervised databases," Expert Syst. Appl., vol. 37, no. 12, pp. 8628–8633, 2010. https://doi.org/10.1016/j.eswa.2010.06.095.

[51] S. Sadiq, Y. Tao, Y. Yan, and M. L. Shyu, "Mining Anomalies in Medicare Big Data Using Patient Rule Induction Method," Proc. - 2017 IEEE 3rd Int. Conf. Multimed. Big Data, BigMM 2017, pp. 185–192, 2017. https://doi.org/10.1109/BigMM.2017.56.

[52] A. Steventon, S. I. Chaudhry, Z. Lin, J. A. Mattera, and H. M. Krumholz, "Assessing the reliability of self-reported weight for the management of heart failure: application of fraud detection methods to a randomized trial of tele monitoring," pp. 1–13, 2017. https://doi.org/10.1186/s12911-017-0426-4.

[53] A. Tagaris, P. Mnimatidis, D. Koutsouris, and S. Member, "Implementation of a Prescription fraud detection software using RDBMS tools and ATC coding" no. November, pp. 5–7, 2009. https://doi.org/10.1109/ITAB.2009.5394458.

[54] D. Thornton, G. Van Capelleveen, M. Poel, J. Van Hillegersberg, and R. M. Mueller, "Outlier-based Health Insurance Fraud Detection for U.S. Medicaid Data," Proc. 16th Int. Conf. Enterp. Inf. Syst., pp. 684–694, 2014.

[55] S. Viaene and G. Dedene, "Insurance fraud: issues and challenges," Geneva Pap. Risk Insur. Pract., vol. 29, no. 2, pp. 313–333, 2004. https://doi.org/10.1111/j.1468-0440.2004.00290.x.

[56] Y. Peng, G. Kou, A. Sabatka, Z. Chen, D. Khazanchi, and Y. Shi, "Application of Clustering Methods to Health Insurance Fraud Detection," Serv. Syst. Serv. Manag. 2006 Int. Conf. IEEE, vol. 1, pp. 116–120, 2006. https://doi.org/10.1109/ICSSSM.2006.320598.

[57] V. C. S. Lee, C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data Minority Report in Fraud Detection: Classification of Skewed Data," no. January 2004.

[58] M. Pejic-Bach, "Profiling intelligent systems applications in fraud detection and prevention: Survey of research articles," ISMS 2010 - UKSim/AMSS 1st Int. Conf. Intell. Syst. Model. Simul., pp. 80–85, 2010. https://doi.org/10.1109/ISMS.2010.26.