International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 8th Aug 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13& issue=issue08

DOI: 10.48047/IJIEMR/V13/ISSUE 08/12

Title Machine Learning Methods for Fraud Detection Techniques in Health Care

Volume 13, ISSUE 08, Pages: 92 -98

Paper Authors

Dr. G.V. Ramesh Babu , Kothapalle Saikumar

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Machine Learning Methods for Fraud Detection Techniques in Health Care

Dr. G.V. Ramesh Babu
Associate Professor, Department of Computer Science, Sri Venkateswara University, Tirupati
gvrameshbabu74@gmail.com
Kothapalle Saikumar
Master of Computer Applications,
Sri Venkateswara University, Tirupati
kothapallesaikumar588@gmail.com

Abstract
Medical services are an essential part in individuals' lives, particularly for the rising older populace, and should be moderate. Federal medical insurance is one such medical care program. Cases extortion is a significant supporter of expanded medical care costs, yet its effect can be diminished through misrepresentation location. In this paper, we contrast a few AI techniques with recognize Medicare misrepresentation. We play out a near report with regulated, solo, and crossover AI approaches utilizing four execution measurements and class unevenness decrease by means of oversampling and a 80-20 undersampling technique. We bunch the 2015 Medicare information into supplier types, with misrepresentation names from the List of Excluded Individuals/Entities data set. Our outcomes show that the fruitful recognition of false suppliers is conceivable, with the 80-20 testing strategy exhibiting the best presentation across the students. Moreover, managed strategies performed better compared to solo or half breed strategies, yet these outcomes changed dependent on the class awkwardness testing method and supplier type.

Indexed Terms: Machine Learning, Fraud, Technique, Health Information, Deep Learning.

Introduction

Medical care is fundamental in individuals' lives and it should be reasonable. The medical care industry is a complex framework with various moving parts. It is extending at a speedy speed. Simultaneously, misrepresentation in this industry is transforming into a basic issue. One of the issues is the abuse of the clinical protection frameworks. Manual recognition of cheats in the medical care industry is exhausting work. [1] As of late, AI and information digging methods are utilized for naturally distinguishing medical services fakes.

Consequently, it is key to have master specialists ready to dissect and treat disorders in different bits of the body. This prompts a couple of kinds of treatment procedures that specialists complete for patients in different specialties. The mark of the prosperity business is to viably fill in whatever number patients as could sensibly be anticipated. Regardless, with every treatment, there is an expense related with each help gave. Specialists, road drug specialists, and clinical staff should be paid for their time and capacity. Tallying diverse clinical accommodations. Generally speaking these expenses are not sensible to the patients. Likewise, assurance plans are used to allocate costs commonly representing patients in the clinical consideration system and pay for the fundamental people and equipment.

Essentially likewise with any insurance system, there is a chance for misuse or distortion works out. [2]

Clinical consideration deception is logically apperceived as one of the veritable social concerns. Clearly, clinical consideration deception is an issue for the public position and there is a prerequisite for more fruitful acknowledgment methodologies. To perceive clinical consideration distortion requires a ton of effort with wide clinical data. [3]

Customarily, medical care misrepresentation location extraordinarily relies upon the experience of space specialists, which is sufficiently mistaken, costly, and tedious. Manual discovery of medical care misrepresentation includes a couple of evaluators who physically survey and recognize the dubious clinical protection claims which require a lot of exertion. However, the cutting-edge advances of AI and information mining strategies prompted more proficient and mechanized recognition of medical service cheats. [4] There has been a developing interest in digging medical care information for extortion location as of late. This paper audits the different methodologies utilized for distinguishing the false exercises in Health protection guarantee information.

Medical care has become a significant consumption in the US since 1980. Both the size of the medical services area and the huge volume of cash included make it an alluring misrepresentation target. As per the Office of Management and Budget, in 2010, about 7%, or around $47.9 billion of the US'S Medicare use was lost due to fraud1. Accordingly, viable misrepresentation recognition is significant for diminishing the expense of the medical care framework. Distinguishing medical care misrepresentation and misuse, be that as it may, needs concentrated clinical information. Numerous medical coverage frameworks depend on human specialists to physically survey protection asserts and recognize dubious ones. [5] This outcomes in both framework advancement and cases to audit is tedious, particularly for the enormous public protection programs in nations like the US. Lately, structures for planning electronic cases have been logically executed to normally perform audits and overviews of cases data. These structures are proposed for recognizing locales requiring exceptional thought, for instance, mistaken or insufficient data input, duplicate cases, and restoratively non-covered organizations. Yet these systems may be used to recognize specific sorts of deception, their coercion disclosure capacities are by and large limited since the area essentially relies upon pre-described fundamental rules dictated by space trained professionals.

Consequently, for generating more successful extended findings, numerous scientists have endeavored to foster more modern enemy of misrepresentation approaches joining information mining, AI, or different strategies. Contrasted with the current extortion identification framework, these recently proposed [6] mechanisms locations were lot of confounded undertakings, such as making the task automated process misrepresentation designs from information, determine the "extortion probability" of each case to focus on some dubious cases, and distinguish another sorts of misrepresentation that were not recently reported. The current proposed medical services extortion identification approaches in the writing can be named three classifications: regulated methodology, for example, choice tree and neural organization, utilized when chronicled misrepresentation information is free and marked; unaided methodology, like grouping, utilized when there is no named recorded extortion [7] information; and crossover approach, which consolidates managed and solo methodologies and typically utilize solo ways to deal with work on the exhibition of the directed methodology.

This paper plans to recognize medical services deceitful conduct, break down the attributes of medical services information, and survey and think about presently proposed extortion recognition approaches utilizing medical services [8] information just as their relating information preprocess, and examine the future exploration bearings. In particular, this paper starts with a foundation information presentation of the US medical services framework and its extortion conduct. Segment 3 investigates the attributes of medical care information utilized or can be utilized in scholarly examination. Then, at that point we audit and think about the as of now proposed misrepresentation location approaches utilizing medical services information in Section 4. In Section 5, we propose a bunching model including geo-area data. Segment 6 examines future examination bearings and makes a few inferences. [9]

What is Machine Learning?

It is a branch of computer science as well as Artificial Intelligence (AI) where it focuses on the working of the algorithms, procedures as well as pseudo codes for performing the task. The main agenda of developing Machine Learning algorithms is to improve the accuracy. IBM has a great place in improving the Machine Learning techniques and mechanisms. [10] Present the techniques like Netflix, various recommendation systems works on the Machine Learning recommendation systems. By machine learning we can perform various classification techniques, regression methods, recommendation techniques etc., by doing so it will improve the development of business. Generally Machine Learning works on various techniques namely: [11]

- By decision process
- Error function
- Model optimization function

The concept of Machine Learning comprises of various methods such as: [12]

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning

As per real world working in the Machine Learning there are many of the use cases. Few of them are: [13]

- Speech recognition.
- Customer service.
- Computer vision.
- Automated stock trading.
- Recommended Engines.

Apart from the more number of benefits in the domain of Machine Learning few of the challenges were leftover for further research in this domain. [14] Many of the researchers, organizations and different funding agencies [15] were working on the bugs in this domain. The major challenges in this domain were:

- Technological similarity.
- Privacy. [16]
- AI impacts on jobs.
- Bias and Discrimination.[17]
- Accountability.

Types of Frauds in Health Care [18]
Based on the occasion or situation the type of frauds will change in health care. The health care fraud happens various ways in different countries. [19] By merging the group of people the fraud is going to happen or takes place. The various frauds are:

- Frauds by service providers [20]
  - Fraud occur in the billing [21]
  - Billing will be done separately apart from treatment this is called Unbundling.
  - They may bill more prices for the products then the actual price. [22]
  - To claim the insurance the service providers may bear other amounts on users.
  - To justify the medicines which are not selling out the shopkeepers will make customers to take.
- Fraud by the insurance agents [23]
  - To get lower premium amount they may give wrong data of the person like age etc.,
  - People may claim on medical services which they didn't needed. [24]

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

  - To get insurance the people uses other people insurance for covering the insurance.
- Frauds done by insurance transmitters [25],[26],[27],
  - Fake re-embossment
  - Misrepresenting the benefit
  - Service statements
- Conspiracy Frauds [28]

In these type of fraud more than one people [29] will involve like for making fraud the patient, doctor or doctor, insurance company will be acted as major role.

Basic needed data for processing

For the defining as well as for the processing [30] of the task we need general entities like which were shown in figure 1.

Personal Attributes	Payment Detail	Settlement Detail	Hospitalized Detail
Person ID	Detail-ID	Settlement-ID	Hospitalized-ID
Sex	Person ID	Person ID	Person ID
Age	Insurance pay	Total cost	Type
Medical insurance usage	Drug name	Type of insurance	Department code
Outpatient amount	Number of medications	Medical insurance costs	Hospitalized days
Hospitalized amount	...	...	...

Figure 1: General Entities needed for processing

Based on the disease for the treatment of the disease is characterized into various sub-divisions such as

- The sample number which changes from one disease to other [31]
- Disease may be caused from the other disease or primary disease, charcoal disease etc.,
- The general count for the disease as well as drug count [32] is equals to twenty thousand. [33]
- Because of different hospitals having various management systems so the drug names as well as disease names will be varies by at least to 11%. [34]
- The losses of data happen because of redundant data, loss of data as well as privacy of data. [35]

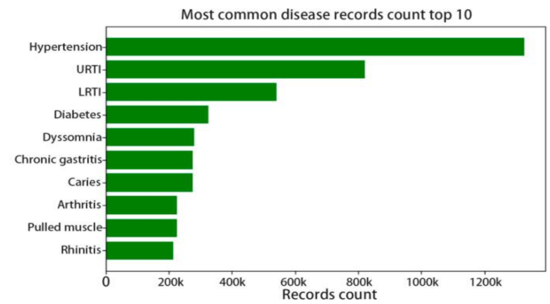

Figure 3: Top disease records in a country

Techniques for Detecting Fraud in Health Care

There are diverse baffling and tangled models with various minimal immaterial subtleties associated with Fraud,[36] whose data is amassed throughout a postponed time frame. It is incredibly difficult to perceive these models in the current events, where we have a gigantic social affair of data and relatively [37] few strategies for evaluating them. Several analysts used to manage a considerable number of [38] clinical consideration claims. Thusly, typically experienced experts would regulate distortion acknowledgment. Regardless, in light of an immense arrangement of data, this strategy becomes drawn-out and inefficient. [39] Upgrades in data mining and AI instruments center around robotized structures for deception revelation. For the area of irregularity and ID of coercion social profiling techniques subject to AI systems are used and therefore, the norm of direct of each person, drawn in with the clinical benefits structure, [40] is intended to notice and check it for any induction from the standards. Data mining techniques are gathered into two arrangements as managed and independent learning by a huge bit of the examiners. Notwithstanding, once in a while, close by these two procedures, semi-coordinated learning is moreover connected with the gathering.

Conclusion

In this paper, medical services extortion, sorts of medical services cheats, types and wellsprings of medical services information, and techniques for medical care fakes were considered. Different examinations are investigated in the writing. It is reasoned that in the medical care industry, 'Information' is a central issue. The significant piece of the information comes from administrative assets [40] and private[41] insurance agencies. Predominantly, AI and information digging are utilized for Healthcare extortion recognition. Directed, solo and semi-regulated learning are the three classes of Machine learning draws near. By and large, semi-administered learning approaches are utilized by numerous specialists. In any case, to distinguish fakes in the medical care framework all the more productively, new semi-regulated learning approaches can be proposed in couple of cases. [42] However, to disguise every one of the examples of medical services extortion, there doesn't exist a specific standard methodology or examples. It very well may be finished up from this audit that the high level AI procedures and recently obtained wellsprings [43] of medical care information would be impending subjects of interest to make medical services reasonable, to work on the adequacy [44] of medical care extortion identification, and to present a top-quality on medical services frameworks.

Summary of the work

Here in this paper we have presented various mechanisms in the designing of various fraud techniques in Health care sector using Machine Learning. Here for doing these tasks various mechanisms were designed such as Bayesian technique, decision tree, Neural Network mechanism etc., were designed for the development of the work. Based on the used or proposed algorithm the efficiency of the particular algorithm is calculated. Here the responsiveness of the algorithm is also designed based on the latency time of the work.

Type	Methods	Fraudulent Behavior
Supervised methods	Neural Network	• Service Providers' Fraud • Insurance Subscribers' Fraud
	Decision Tree	• Service Providers' Fraud • Insurance Subscribers' Fraud
	Genetic Algorithm & KNN	Service Providers' Fraud
	Rule-based Classifier & BN	Insurance Subscribers' Fraud
Unsupervised methods	SOM	Service Providers' Fraud
	Association Rules	Insurance Subscribers' Fraud
	Rule-based Method	Service Providers' Fraud
	Finite Mixture Model	Insurance Subscribers' Fraud
	Clustering	Service Providers' Fraud
	Subjective Utility model	Insurance Subscribers' Fraud
Hybrid methods	SOM & Neural Network	Service Providers' Fraud
	Clustering & Decision Tree	Insurance Subscribers' Fraud

Table 1: Overall brief regarding various algorithms

As per the data collected from the above table we can conclude that the supervised data has more efficient for performing various tasks. However supervised data technique is very efficient for determining the task. The result generation and obtaining of the values were very tough. For considering the accuracy as well as finding the correctness of the information is very good.

References

[1] Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113.
[2] Behdad, Mohammad, et al. "Nature-inspired techniques in the context of fraud detection." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 (2012): 1273-1290.
[3] Konasani, Venkatareddy, Mukul Biswas, and Praveen Krishnan Koleth. "Healthcare fraud management using big data analytics." An Unpublished Report by Trendwise Analytics, Bangalore, India (2012).
[4] National Health Care Anti-Fraud Association. "Health Care Fraud–A Serious and Costly Reality For All Americans." April 2005 (2007).
[5] Yang, Wan-Shiou. "A Process Pattern Mining Framework for the Detection of Health Care Fraud and Abuse." National Sun Yat-Sen University, Taiwan (2003).
[6] Liu, Qi, and Miklos Vasarhelyi. "Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information." In 29th World Continuous

Auditing and Reporting Symposium (29WCARS), Brisbane, Australia. 2013.
[7] Thornton, Dallas, Roland M. Mueller, Paulus Schoutsen, and Jos van Hillegersberg. "Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection." Procedia technology 9 (2013): 1252-1264.
[8] Bauder, Richard A., Taghi M. Khoshgoftaar, Aaron Richter, and Matthew Herland. "Predicting medical provider specialties to detect anomalous insurance claims." In Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on, pp. 784-790. IEEE, 2016.
[9] Branting, L. Karl, Flo Reeder, Jeffrey Gold, and Timothy Champney. "Graph analytics for healthcare fraud risk estimation." In Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on, pp. 845-851. IEEE, 2016.
[10] Musal, Rasim Muzaffer. "Two models to investigate Medicare fraud within unsupervised databases." Expert Systems with Applications 37, no. 12 (2010): 8628-8633.
[11] Copeland, Leanndra, Dana Edberg, Anna K. Panorska, and Jeanne Wendel. "Applying business intelligence concepts to Medicaid claim fraud detection." Journal of Information Systems Applied Research 5, no. 1 (2012): 51.
[12] Bauder, Richard A., and Taghi M. Khoshgoftaar. "A probabilistic programming approach for outlier detection in healthcare claims." In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on, pp. 347-354. IEEE, 2016.
[13] Bauder, Richard A., and Taghi M. Khoshgoftaar. "A novel method for fraudulent medicare claims detection from expected payment deviations (application paper)." In Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on, pp. 11-19. IEEE, 2016.
[14] van Capelleveen, Guido, Mannes Poel, Roland M. Mueller, Dallas Thornton, and Jos van Hillegersberg. "Outlier detection in healthcare fraud: A case study in the Medicaid dental domain." International journal of accounting information systems 21 (2016): 18-31.
[15] Rudman, William J., John S. Eberhardt, William Pierce, and Susan Hart-Hester. "Healthcare fraud and abuse." Perspectives in Health Information Management/AHIMA, American Health Information Management Association 6, no. Fall (2009).
[16] Joudaki, Hossein, Arash Rashidian, Behrouz Minaei-Bidgoli, Mahmood Mahmoodi, Bijan Geraili, Mahdi Nasiri, and Mohammad Arab. "Using data mining to detect health care fraud and abuse: a review of literature." Global journal of health science 7, no. 1 (2015): 194.
[17] Jyothsna, V., VV Rama Prasad, and K. Munivara Prasad. "A review of anomaly based intrusion detection systems." International Journal of Computer Applications 28, no. 7 (2011): 26-35.
[18] Li, Jing, Kuei-Ying Huang, Jionghua Jin, and Jianjun Shi. "A survey on statistical methods for health care fraud detection." Health care management science 11, no. 3 (2008): 275 - 287.
[19] Srinivasan, Uma, and Bavani Arunasalam. "Leveraging big data analytics to reduce healthcare costs." IT professional 15, no. 6 (2013): 21-28.
[20] Feldman, Keith, and Nitesh V. Chawla. "Does medical school training relate to practice? Evidence from big data." Big data3, no. 2 (2015): 103-113.
[21] Ko, Joan S., Heather Chalfin, Bruce J. Trock, Zhaoyong Feng, Elizabeth Humphreys, Sung-Woo Park, H. Ballentine Carter, Kevin D. Frick, and Misop Han. "Variability in Medicare utilization and payment among urologists." Urology 85, no. 5 (2015): 1045-1051.
[22] Abbott DW, Matkovsky IP, Elder JF (1998) An evaluation of high-end data mining tools for fraud detection. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA
[23] Bonchi F, Giannotti F, Mainetto G, Pedreschi D (1999) A classification-based methodology for planning auditing strategies in fraud detection. In Proceedings of SIGKDD99, 175–184.
[24] Chan CL, Lan CH (2001) A data mining technique combining fuzzy sets theory and Bayesian classifier—an application of

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

auditing the health insurance fee. In Proceedings of the International Conference on Artificial Intelligence, 402–408.
[25] Cox E (1995) A fuzzy system for detecting anomalous behaviors in healthcare provider claims. In: Goonatilake S, Treleaven P (eds) Intelligent systems for finance and business. Wiley, New York, pp 111–134.
[26] Ghosh S, Reilly D (1994) Credit card fraud detection with a neural network. Proceedings of 27th Hawaii International Conference on Systems Science 3:621–630.
[27] He H, Hawkins S, Graco W, Yao X (2000) Application of Genetic Algorithms and k-Nearest Neighbour method in real world medical fraud detection problem. Journal of Advanced Computational Intelligence and Intelligent Informatics 4(2):130–137.
[28] Ormerod T, Morley N, Ball L, Langley C, Spenser C (2003) Using ethnography to design a Mass Detection Tool (MDT) for the early discovery of insurance fraud. In Proceedings of the ACM CHI Conference.
[29] Ortega PA, Figueroa CJ, Ruz GA (2006) A medical claim fraud/abuse detection system based on data mining: a case study in Chile. In Proceedings of International Conference on Data Mining, Las Vegas, Nevada, USA.
[30] Sokol L, Garcia B, West M, Rodriguez J, Johnson K (2001) Precursory steps to mining HCFA health care claims. In Proceedings of the 34th Hawaii International Conference on System Sciences.
[31] Viveros MS, Nearhos JP, Rothman MJ (1996) Applying data mining techniques to a health insurance information system. In Proceedings of the 22nd VLDB Conference, Mumbai, India, 286–294.
[32] Wei CP, Hwang SY, Yang WS (2000) Mining frequent temporal patterns in process databases. Proceedings of international workshop on information technologies and systems, Australia, 175–180.
[33] Yamanishi K, Takeuchi J, Williams G, Milne P (2004) On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Mining and Knowledge Discovery 8:275–300.
[34] Yang WS (2003) A Process Pattern Mining Framework for the Detection of Health Care Fraud and Abuse, Ph.D. thesis, National Sun Yat-Sen University, Taiwan.
[35] Yang WS (2002) Process analyzer and its application on medical care. In Proceedings of 23rd International Conference on Information Systems (ICIS02), Spain
[36] Yang WS, Hwang SY (2006) A process-mining framework for the detection of healthcare fraud and abuse. Expert Syst Appl 31:56–68.
[37] Williams G, Huang Z (1997) Mining the knowledge mine: The Hot Spots methodology for mining large real world databases. Lect Notes Comput Sci 1342:340–348.
[38] Williams G (1999) Evolutionary Hot Spots data mining: an architecture for exploring for interesting discoveries. Lect Notes Comput Sci 1574:184–193.
[40] Viaene S, Derrig R, Dedene G (2004) A case study of applying boosting Naive Bayes to claim fraud diagnosis. IEEE Trans Knowl Data Eng 16(5):612–620.
[41] Spirtes P, Glymour C, Scheines R (1993) Causation, Prediction and Search. Springer, New York.
[42] Shapiro AF (2002) The merging of neural networks, fuzzy logic, and genetic algorithms. Insurance: Mathematics and Economics 31:115–131.
[43] Ireson CL (1997) Critical pathways: effectiveness in achieving patient outcomes. J Nurs Adm 27(6):16–23.
[44] Phua C, Alahakoon D, Lee V (2004) Minority report in fraud detection: classification of skewed data. SIGKEE Explorations 6(1):50–59.