HEALTH INSURANCE FRAUD DETECTION

Co-authored by

Jasmine Kaur Gill and Shaun Aghili

Information Systems Assurance Management

Concordia University of Edmonton

Project Report

Submitted to the Faculty of Graduate Studies,

Concordia University of Edmonton

In Partial Fulfillment of the

Requirements for the Final

Research Project for the Degree

MASTER OF INFORMATION SYSTEMS ASSURANCE MANAGEMENT

Concordia University of Edmonton

FACULTY OF GRADUATE. STUDIES

Edmonton, Alberta

December 2020

HEALTH INSURANCE FRAUD DETECTION

Jasmine Kaur Gill

Approved:

Shaun Aghili [Original Approval on File]

Shaun Aghili Date: December 14, 2020

Primary Supervisor

Edgar Schmidt [Original Approval on File]

Edgar Schmidt, DSocSci Date: December 15, 2020

Dean, Faculty of Graduate Studies

Table of Contents

List of Figures ..... *v*

List of Tables ..... *v*

Abstract ..... *vi*

Introduction *1*

Background **1**

Literature Review *3*

Healthcare Insurance Fraud **3**

Impact of Healthcare Insurance Fraud **4**

Healthcare Privacy Legislation **6**

Types of Healthcare Insurance Fraud **7**

Electronic Claim Management System for Healthcare Insurance **11**

Using Intelligent Fraud Detection Techniques **14**

Data Mining for Healthcare Insurance Fraud Detection **15**

Supervised Learning. 16

Unsupervised Learning. 17

Hybrid Learning. 18

Use of Data mining algorithms in Health Insurance Fraud Detection **18**

Challenges in Health Insurance Fraud Detection **20**

Research Methodology *21*

Presentation of Results *24*

SAS Viya: Detection and Investigation for Healthcare **24**

H2O.ai: Driverless AI **25**

IBM Counter Fraud Management **26**

Recommendations and Conclusions *29*

Bibliography *31*

List of Figures

Figure 1. Categories of Health Insurance Fraud (Liu & Vasarhelyi, 2013) 8

Figure 2. Categories of healthcare fraud: Service Providing Pattern and Service Availing Patterns (Matloob, Khan, Rahman, & Hussain, 2020) 9

Figure 3. The Electronic Claim Management System (Kose, Gokturk, & Kilic, 2015)... 13

List of Tables

Table 1. Advantages & Disadvantages of Supervised & Unsupervised Learning 17

Table 2. Healthcare fraud detection methods (Liu & Vasarhelyi, 2013)..... 19

Table 3. Data mining methods used in healthcare fraud detection domain (Joudaki, et al., 2015) 20

Abstract

Healthcare Fraud is an act that causes a loss of billions of dollars every year across the world. The impact of healthcare fraud is very far-reaching and affects the efficiency of healthcare systems. To thwart the perpetrators early on, intelligent fraud detection technologies are required to predict the strategies and schemes in sophisticated organized fraud cases. The work already done in the field of healthcare insurance fraud has been discussed in a systematic literature review with a particular focus on its detection techniques and various aspects associated with it. There are many application solutions available in the market that can be used to detect healthcare fraud. This paper provides a side-by-side comparison of three existing application solutions for fraud detection pertaining to the healthcare insurance domain. The comparative study forms the basis for listing the features desired in an ideal healthcare fraud detection application solution. The research done in this paper aimed to provide a list of features of a best-in-class healthcare insurance fraud detection application solution. Also, some of the key challenges and issues faced by the fraud detection industry have been highlighted along with the potential future direction of research.

Keywords Healthcare Fraud, Fraud Detection, Data Mining, Machine Learning, Artificial Intelligence

Introduction

Background

Fraud is a major threat to organizations of all types and sizes throughout the world (ACFE, 2018). In fact, recent years have witnessed an exceptional surge in the number of incidents of fraud, mainly owing to the advancements in technology making fraud an important area for exploration (Kou, Lu, Sirwongwattana, & Huang, 2004).

The Office of Public Health at the U.S. Department of Justice recently released a report regarding a healthcare fraud takedown which had caused approximately 6 billion USD of losses (Office of Public Health, 2020). The report by the US Department of Justice described that the case involved 345 charged defendants across the US which included a number of licensed health professionals. The defendants were charged with the submission of at least 6 billion USD in false claims to the private and public health care programs.

The case described above was termed as “historic in size and scope” by the Acting Assistant Attorney General Brian C. Rabitt. The case portrayed the medical professionals who were held responsible for the exploitation of the healthcare benefits plans and the patients’ private information for personal gain. The false claims submitted included 4.5 billion USD worth of claims related to telemedicine, around 845 million USD worth substance abuse treatment facilities, also known as “Sober Homes”, and also, about 806 million USD in illegal opioid distribution schemes across the US.

The Health Insurance Industry has become a common target for fraudsters given the enormous amount of money involved (Liu & Vasarhelyi, 2013). According to the Canadian Institute of Health Information (CIHI), the total spending in Canadian

healthcare was estimated to be around $264 billion in the year 2019. As per the industry standards, 2% to 10 % of the total healthcare spending, an estimated total of 5.2-26 billion CAD is lost every year as a result of health insurance fraud (BlueCross, 2016). Therefore, health insurance fraud, is a significant problem in the development of a cost-effective healthcare system (Liu & Vasarhelyi, 2013).

As depicted in the case history and the problem statement above, Fraud is one of the biggest challenges faced by the health insurance industry and proves to be very costly (Rawte & Anuradha, 2015). Health insurance fraud refers to the deliberate deception of an insurance company resulting in illegitimate payment of healthcare benefits to an individual or group of individuals (Rawte & Anuradha, 2015). Healthcare fraud has become a pressing problem which causes significant losses to the industry and also results in growing costs of healthcare programs. For governments throughout the world, elimination of fraud is proving to be a major issue (Evbayiro, 2011). Intelligent detection methods provide a way of combating fraud by providing organizations with the ability to protect themselves against the ever-evolving fraud methods and schemes (West & Bhattacharya, 2016).

The purpose of this research project was to propose features of a conceptual fraud detection approach in order to better detect the instances of fraud in the North American Health Insurance Industry. Furthermore, this research deliverable also includes a side-by-side comparison and analysis of the major, existing fraud detection technologies predominantly being used in the North American Region.

The organization of this research paper is as follows: The introduction section defines what constitutes healthcare fraud and explains the need for effective and efficient

detection methods. The literature review section provides an insight into the work already done in the field of healthcare fraud detection. The literature review section is followed by a methodology section which includes the research scope, limitations, questions and research deliverable procedure. The following section will include a discussion on characteristics of three application solutions along with a side-by-side comparison of the application solutions. A list of desired key features of a more effective healthcare fraud detection technique will also be provided. The last section will include the concluding statements, contribution to knowledge and future scope of the research done.

Literature Review

The following section discusses healthcare insurance fraud, its impact and its types. Also, this section provides a brief discussion on electronic claims management system and various data mining methods being deployed for healthcare insurance fraud detection.

Healthcare Insurance Fraud

Rebecca S. Busch in her book, *Healthcare Fraud: Auditing and Detection Guide,* describes Healthcare Fraud as "a knowing and intentional execution of a scheme to defraud a healthcare benefit program". Healthcare Fraud has been defined by the National Healthcare Anti-Fraud Association (NHCAA) as "an intentional deception or misrepresentation that an individual or entity makes knowing that the misrepresentation could result in some unauthorized benefit to the individual, or to the entity or to some other party" (BlueCross, 2016).

Thornton et al. (2015) characterizes Healthcare Fraud as a crime which is ever evolving, where new schemes emerge on a regular basis (Thornton, Brinkhuis, Amrit, &

Aly, 2015). Fraudsters, as the technology is advancing, are also becoming increasingly innovative in their methods for perpetrating fraudulent schemes (West & Bhattacharya, 2016).

In order to make a continual progress in the improvement of the healthcare industry, it is very important to understand the working of the fraudsters and the methods that are used by the fraudsters (Thornton, Brinkhuis, Amrit, & Aly, 2015).

FBI (2006) states that it is very difficult to place an exact value on the theft that is done through the means of insurance fraud. Fraud is supposed to be deliberately undetectable in nature (FBI, 2006). The complexity and confusing nature of the healthcare system makes it very hard to uncover fraudulent activities (Abdallah, Maarof, & Zainal, 2016). So, the number of instances of fraud that are discovered is much lower than the actual number of fraud instances (FBI, 2006). Fraudulent transactions are masqueraded as legitimate transactions so there will always be room for improvement in fraud detection domain (Wright, 2015).

Impact of Healthcare Insurance Fraud

Healthcare Systems have become an important part of modern life (Abdallah, Maarof, & Zainal, 2016). According to the Canadian Institute of Health Information (CIHI), the total healthcare spending in Canada was estimated to be around $264 billion in the year 2019. As a result, the healthcare industry has become a target for fraudsters (Abdallah, Maarof, & Zainal, 2016). As per the industry standards, 2% to 10 % of the total healthcare spending, an estimation of 5.2 billion to 26 billion CAD (BlueCross, 2016) is lost every year as a result of medical claims fraud. Therefore, healthcare

insurance fraud has a wide-reaching and significant impact (Abdallah, Maarof, & Zainal, 2016).

Van Capelleveen et al. (2016) also describes healthcare fraud as a pressing problem which causes significant and growing costs of healthcare programs. Both Van Capelleveen et al. (2016) and Abdallah et al. (2016) have described the healthcare system as a very complex system with many parties involved. Healthcare systems also involve many rules and regulations which makes the discovery of fraudulent incidents very hard (Abdallah, Maarof, & Zainal, 2016).

Stowell et al. (2018) described three parties who stand to suffer financially as an impact of Healthcare Fraud: Insurance Holders, Taxpayers, and Businesses. Insurance Holders suffer by having to pay higher premiums for insurances, reduced benefits and coverage. In public health plans, taxpayers have to pay more in order to cover the healthcare expenditures as a result of rising fraud in healthcare industry. Businesses suffer by experiencing an overall increased cost of doing business by having to pay extra amounts for employee health insurance as a by-product of increasing medical insurance fraud (Stowell, Schmidt, & Wadlinger, 2018).

National Healthcare Anti-Fraud Association's article, "*The Challenge of Healthcare Fraud",* describes the physical risk to patients as a very dangerous by-product of healthcare fraud. NHCAA (2018) exemplified the physical risk the patients are subjected to, using two incidents of performing excessive, non-required surgeries on patients (NHCAA, 2018).

An article by Alberta BlueCross also states that health insurance fraud impacts everybody including the ones who are not at fault. The effects of healthcare insurance

fraud can be (Alberta BlueCross, n.d.): hike in the benefit costs and premiums, reduction in coverage under benefits plan, denial of reimbursement of claims, and creation of false records for a plan member.

Healthcare Privacy Legislation

According to *Healthcare and Abuse Control Program Report* by Department of Health and Human Services, Health Insurance Portability and Accountability Act (HIPAA) consolidated and strengthened the efforts being made to combat fraud in the health insurance domain. The HIPAA provided a comprehensive program to manage fraud against both public and private health insurance plans (HHS, 2018).

Hyman (2002) provides a detailed discussion on the relation of the Health Insurance Portability and Accountability Act (1996) and healthcare fraud. This paper discusses about the statutory and administrative framework which existed for healthcare fraud control before the HIPAA was passed in the year 1996 in the US. The author also discusses how HIPAA changed the pre-existing control framework for fraud. HIPAA created strict laws for healthcare fraud (Hyman, 2002). HIPAA created three programs for to pursue fraud cases: The Control Program, the Integrity program, and the Beneficiary program (Hyman, 2002). HIPAA implemented continuous stream of funds for fraud control as well (Hyman, 2002).

D'Agostino & Woodward (2010) discussed the key role of Electronic Health Records in modernizing the Canadian Healthcare System. The authors suggest that accurate and accessible health information is necessary for informed decision making. Various federal and provincial privacy laws exist in Canada that regulate the protection of patients' personal health information which includes the establishment of various

standards for patient privacy rights (D'Agostino & Woodward, 2010). The authors mention the existence of freedom of information and also, the protection of privacy statutes in Canada to protect the personal health information in custody of public bodies such as government hospitals and regional health authorities. Personal Information Protection and Electronic Document Act (PIPEDA) was enforced in the year 2001 but its application was extended to healthcare data in 2004 to regulate the health information being handled in the private sector (D'Agostino & Woodward, 2010).

In every Canadian province, a public sector legislation exists which applies to provincial government entities and takes precedence (Seth, n.d.). Some of them are (Seth, n.d.):

- Personal Information Protection Act (Alberta, 2004)
- Personal Information Protection Act (British Columbia, 2004)
- Act Respecting the Protection of Personal Information in the Private Sector (Quebec, 1994)
- Personal Health Information Protection Act (Ontario, 2004)
- Personal Health Information Privacy and Access Act (New Brunswick, 2009)
- Personal Health Information Act (Newfoundland & Labrador, 2011)

Types of Healthcare Insurance Fraud

Eighteen healthcare insurance fraud type have been identified by Thornton et al. (2015) in the paper, *Categorizing and Describing the Types of Fraud in Healthcare*. FBI's Financial Crimes Report to the Public identifies eight healthcare insurance fraud types (FBI, 2006). Rawte & Anuradha (2015) discussed five healthcare insurance fraud types in their paper describing health insurance fraud detection with the use of Data

Mining techniques. An article by the National Healthcare Anti-Fraud Association (2018) also mentioned the nine common healthcare insurance fraud types. ACFE's article by Charles Piper mentions ten popular health insurance fraud schemes.

Stowell et al. (2017) in the paper, *Healthcare fraud under the microscope: improving its prevention,* has discussed twelve types of healthcare fraud schemes demonstrating each one with the example of a case.

Liu & Vasarhelyi (2013) discussed the various healthcare fraud types by classifying them into three different categories. This paper also discusses conspiracy fraud which is basically patients colluding with the physicians. The authors also mention that the service provider's fraud accounts for the highest amount among all types of healthcare fraud. Li et al. (2007) also classified various health care fraud schemes into three types: Service Providers' Fraud, Insurance Subscribers' Fraud and, Insurance Carriers' Fraud.

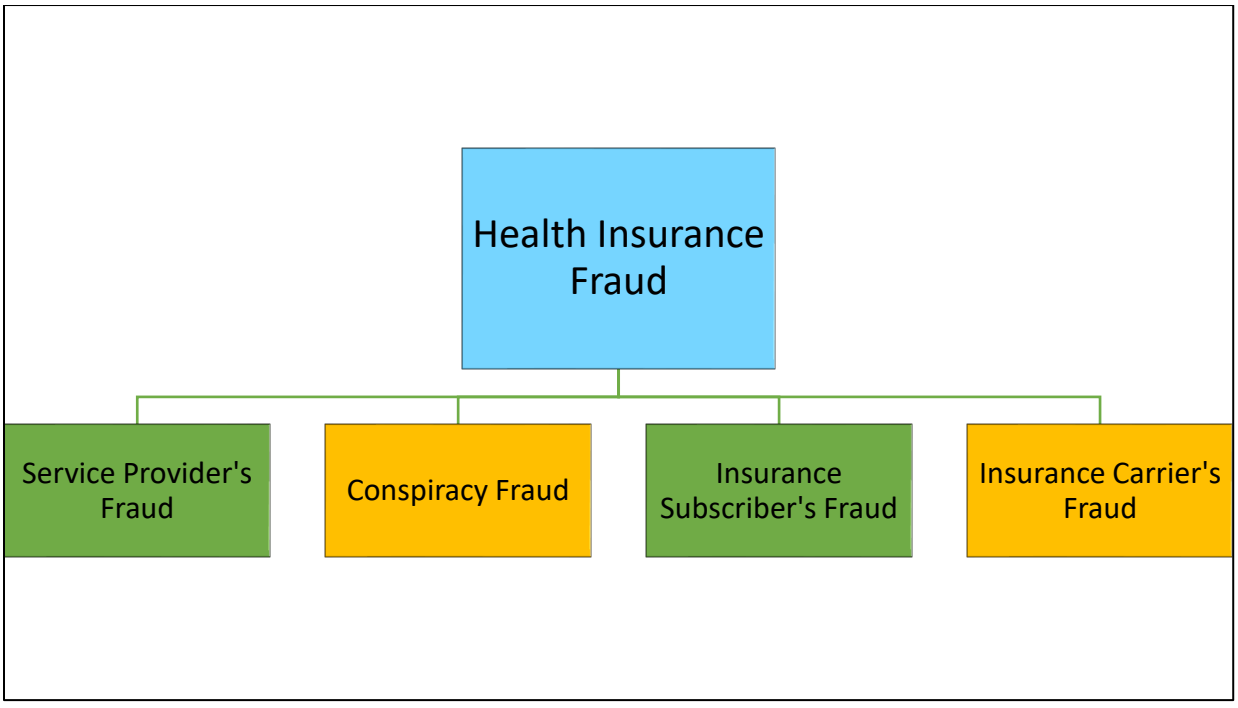

Figure 1. Categories of Health Insurance Fraud (Liu & Vasarhelyi, 2013)

Matloob et al. (2020) explained healthcare fraud as classified into two kinds: service-availing patterns and service-providing patterns. The service availing patterns basically include the services that are being taken by the patients. Service providing patterns is basically misrepresentation by the medical professionals involved, that is, doctors, pharmacies or hospitals itself. The following figure illustrates the concept of both service providing and availing patterns as described in the paper.

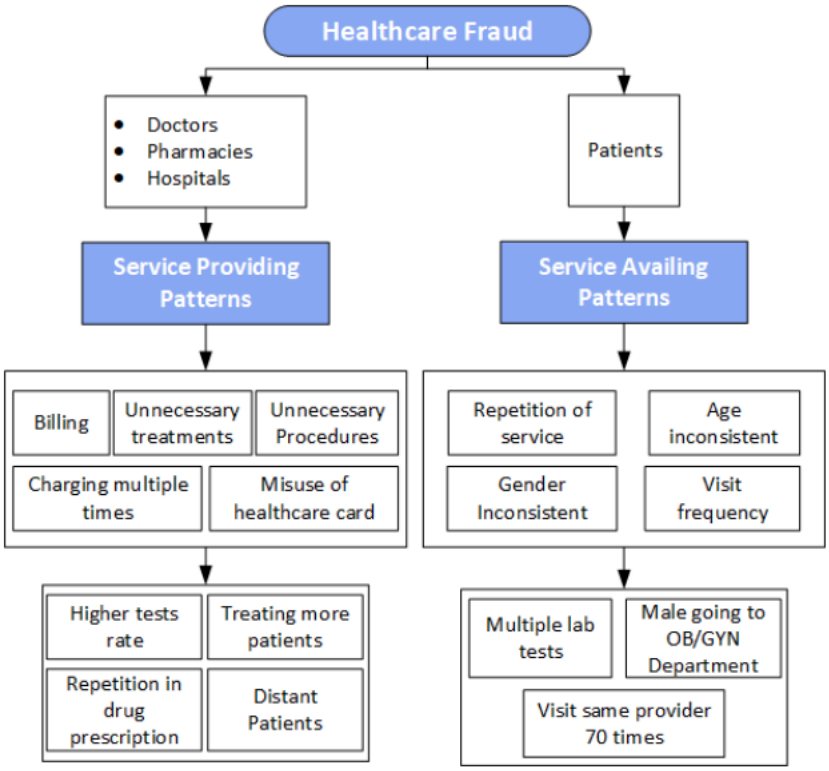

Figure 2. Categories of healthcare fraud: Service Providing Pattern and Service Availing Patterns (Matloob, Khan, Rahman, & Hussain, 2020)

Some of the common healthcare insurance fraud types have been discussed below:

Upcoding. This is one of the most prevalent schemes in the area of medical insurance fraud. Upcoding is basically billing for a service which is costlier than the one which has actually been rendered (Bauder, Khoshgoftaar, & Seliya, 2017). A report by the Federal

Bureau of Investigation describes Upcoding as a billing practice in which the receipt submitted by a health care provider using a procedure code that results in a higher payment than the actual code of the service performed (FBI, 2006). Upcoding scheme can apply to either a service or an item.

Duplicating Claims. Duplicating claims, as the name suggests, refers to the fraudulent practice of submitting multiple claims for the same service or procedure (Thornton, Brinkhuis, Amrit, & Aly, 2015). Healthcare providers try to bill multiple times for the same service, so as to be paid several times. More insight has been provided by FBI (2006) that that the same exact claim is not filed again. Instead, the healthcare provider changes some part of the claim in order to deceive the health insurer.

Billing for Services not rendered. One of the most popular schemes, where a healthcare provider submits a claim to an insurance provider for a service which was never provided, proved by the lack of supporting documentation (Piper, 2013). This scheme is also known as Phantom Billing. (Thornton, Brinkhuis, Amrit, & Aly, 2015)

Kickbacks. Kickback refers to when a healthcare provider offers to pay or accepts money or something valuable in exchange of referring a healthcare service which is covered by insurance (FBI, 2006). Stowell et el. describes kickbacks as particularly "malicious" as this practice can compromise a health provider's ability of decision-making and makes profit from the scheme as the provider's primary goal rather than the patient's welfare (Stowell, Schmidt, & Wadlinger, 2018). Kickbacks may result in inappropriate medical care, which can include incorrect hospitalization, surgeries, tests, prescription and or use of wrong medication and equipment (Stowell, Schmidt, & Wadlinger, 2018).

Unbundling. Unbundling refers to the reporting of excessive number of claims separately which should be billed together as one service (Abdallah, Maarof, & Zainal, 2016). Unbundling has been exemplified in an article on the FBI website as follows: a hysterectomy cost $1,300 on an average but in the case of unbundling this procedure, health care provider may decide to charge additional $950 for removing ovaries and fallopian tubes, $671 for exploring the abdomen, $250 for performing an appendectomy and $550 for lysis of adhesions. The total would come up to $3,721 (FBI, 2006).

Identity Fraud. Theft of medical entity has been described by Stowell et el. (2018) as a fast-growing area of medical insurance fraud. False or stolen medical identity is used to submit claims against the victim's insurance plan to get the claimed amount and earn profit (Stowell, Schmidt, & Wadlinger, 2018) . Thornton et el. describes Identity fraud in health insurance as an instance when an uninsured person poses as an insured person in order to get a medical service or also, to conceal a certain illness (Thornton, Brinkhuis, Amrit, & Aly, 2015).

Excessive Services. This scheme refers to providing unnecessary services to a patient than what is actually needed (FBI, 2006). Thornton et al. (2015) adds that in the excessive services scheme, documents are sometimes forged in order to justify the payments made. In some cases, unnecessary care is provided when the care providers use treatments which are not proven to work (Thornton, Brinkhuis, Amrit, & Aly, 2015).

Electronic Claim Management System for Healthcare Insurance

Matloob et al. (2020) mentioned the use of "Special Investigation Departments" by most insurers in order to control the fraud in the reimbursement of healthcare bills. In case of a fraudulent payment being detected, a recovery process is initiated, and

appropriate controls are then introduced to avoid any future payment of similar misrepresented billing. The claims case, when successfully identified as fraudulent, is then recognized as an identified fraudulent pattern. These identified fraudulent patterns are used in readjusting the existing system's billing process to prevent future fraudulent occurrences. But this approach can prove to not be so cost-efficient due to the presence of huge variations in the practices and billing patterns of the service providers (Matloob, Khan, Rahman, & Hussain, 2020).

Harrison & Lee (2006) mentioned in their article that electronics claim management system is one of the most widely adopted e-Health technologies in healthcare systems. Liu & Vasarhelyi (2013) also stated electronic claims management systems are increasingly being used for automatic claims processing, auditing and reviewing.

Kose et al. (2015) have described the role of Electronic Claim Processing (ECP) systems in health insurance claims processing. Some of the pros of electronic claim processing systems as described by AMA & CSMS (2013) are: reduces the time and resources used for manual claim processing, identification of claim issues and their resolutions before processing to payer, tracking of claims' progress, and confirmation of a payer's claim receipt with the use of electronic reporting.

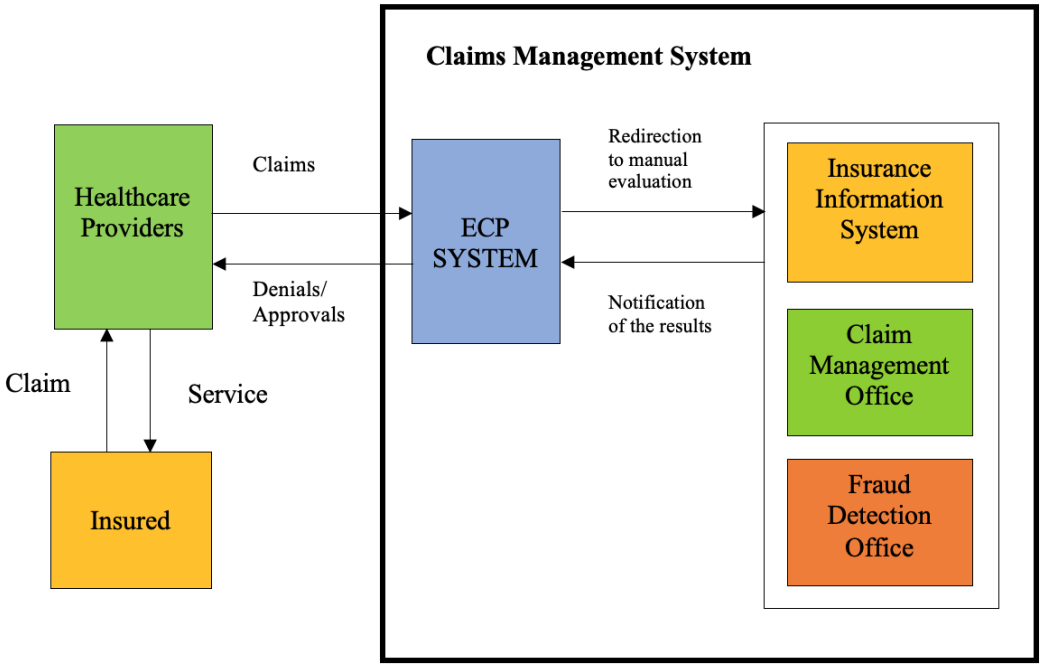

Figure 3. The Electronic Claim Management System (Kose, Gokturk, & Kilic, 2015)

Kose et al. (2015) mentioned the four entities in a claim management system: insurance company, beneficiary, service provider and claim. The claim management system uses electronic claim data which is provided by the healthcare provider (Placeholder1). The ECP system generally has two rule engines: medical rules and insurance rules (Placeholder1). Kose et al. (2015) also describes what both medical and insurance rules indicate. The claim management system also has a fraud detection office which is used to handle cases that cannot be managed by the ECP system due to its shortcomings (Placeholder1).

Both public and private insurers are known to use automated detection systems and human experts to identify and investigate fraudulent claims (Ai, Lieberthal, Smith, & Wojciechowski, 2018).

Using Intelligent Fraud Detection Techniques

Electronic Health Records (EHRs) are being adopted by healthcare and health insuring agencies which has led to the collection of large amounts of healthcare-related data (Bauder, Khoshgoftaar, & Seliya, 2017). In the conventional methods of healthcare fraud detection, few auditors were responsible for handling thousands of submitted claims which made it very difficult to focus on the overall outlook of a provider's behavior. The traditional way is considered time-consuming and inefficient (Joudaki, et al., 2015). Thus, a need of using data analytics solutions has emerged in the sector of healthcare information systems (Bauder, Khoshgoftaar, & Seliya, 2017). EHRs and the growing use of Information Technology has made way for emerging opportunities for more efficient methods of fraud detection.

Liu et al. (2016) also discusses the submissions of hundreds of millions of claims each year to a government or private health-care insurance program. Liu et al. (2016) also suggests that a health insurance fraud detection system is supposed to be well-equipped to deal with the volume and high degrees of diversity present in healthcare data (Liu J. , Bier, Guerra-Gomez, Honda, & al., 2016).

Advancements in machine learning and artificial intelligence have provided with new automated techniques of fraud detection (Joudaki, et al., 2015). The past decades have witnessed the development of many IT tools, with underlying technology such as Data Mining (Segal, 2016). Data Mining and Regression have been identified as the two

dominant approaches in health care fraud detection at claims, provider and facility level analysis (Ai, Lieberthal, Smith, & Wojciechowski, 2018).

Abdallah et al. (2016) briefly mentioned three major advantages of using data mining techniques to detect fraud instances in health insurance. The advantages of data mining techniques are automatic identification of fraud patterns in the provided data, determination of likelihood of occurrence of fraud in each case, and discovery of new fraud types (Abdallah, Maarof, & Zainal, 2016). Data Mining is a very effective and efficient approach being used to combat health insurance fraud (Joudaki, et al., 2015). West & Bhattacharya (2016) also suggested data mining to be very useful in fraud detection area.

Data Mining for Healthcare Insurance Fraud Detection

Segal (2016) describes data mining techniques as the capability to investigate large populations of data for the purpose of fraud detection. Data mining provides helpful responses that are easily understood by the user (Segal, 2016). Data Mining tools have the ability to detect behaviour anomalies of users by comparing against already known models and profiles. Segal (2016) also enlisted three most popular data mining programs along with key features: IBM SPSS Modeler, SAS Data Mining and, RapidMiner.

Jain et al. (2019) described a detailed methodology of data mining as used in fraud detection. Fraud detection follows the basic flow: feature selection, data representation, data collection, data preprocessing, data mining, post-processing and performance evaluation. It was found that the neural networks, logistic model, bayesian belief networks and decision trees were the most widely used methods in fraud detection. But

out of the forty-one studies reviewed by the authors, none of the studies used either of the methods for healthcare insurance fraud detection.

Bauder et al. (2017) reviewed the various facets of healthcare fraud analysis and detection mainly focusing on upcoding fraud. Bauder et al. (2017) also described the data used in data mining can be either structured or unstructured. Structured data is organized data which makes it possible for the data to be stored in a conventional database usually in a tabular format, whereas unstructured data is unorganized in nature. Structured data can be easily processed by data mining techniques in contrast to unstructured data which needs further analysis and parsing methods (Bauder, Khoshgoftaar, & Seliya, 2017).

Abdallah et al. (2016) classified data mining techniques into three types: Supervised, Unsupervised and Semi-supervised. Both Liu et al. (2013) and Bauder et al. (2017) categorized data mining techniques into three types: supervised, unsupervised and hybrid learning. Jing Ai (2008) described the various supervised, unsupervised, and hybrid learning methods used in fraud detection. Rawte & Anuradha (2015) described supervised and unsupervised learning along with their advantages and disadvantages over each other. Supervised learning and unsupervised learning are the most accepted classifications used for categorizing data mining methods (Joudaki, et al., 2015).

Supervised Learning. Supervised Learning (also known as Predictive Learning) is described as the most commonly used learning technique to train the model using pre-defined class labels: legitimate and fraudulent (Rawte & Anuradha, 2015; Abdallah, Maarof, & Zainal, 2016). Training dataset is used to build the model against which the new claims are compared for class prediction. Rawte & Anuradha (2015) also discussed the advantages and disadvantages of supervised learning method for data mining.

Joudaki et al. (2015) added that supervised learning needs the correct categorization of the input data for successful detection of fraud instances. To be able to reflect the new types of fraud models are updated on a regular basis (Joudaki, et al., 2015). Fraud related labelled data is rarely found in publicly available data for supervised learning method (Bauder, Khoshgoftaar, & Seliya, 2017).

Unsupervised Learning. Unsupervised Learning (also known as Descriptive Learning) is described by Rawte & Anuradha (2015) as having no class labels. Unsupervised learning can find the cases with unusual behaviour and can uncover old and new fraud types (Rawte & Anuradha, 2015).

Joudaki et al. (2015) adds that unsupervised learning method assesses the characteristics of a claim in comparison to other claims to find their relation or how the claims differ from each other. Liu & Vasarhelyi (2013) have also discussed the applicability and use of supervised and unsupervised learning methods in healthcare fraud detection. It was concluded in the paper that taking data availability and accuracy into consideration, that unsupervised methods were more applicable in healthcare insurance fraud detection domain.

The following table provides a comparison of advantages and disadvantages of supervised and unsupervised learning (Rawte & Anuradha, 2015):

Table 1. Advantages & Disadvantages of Supervised & Unsupervised Learning

	Supervised Learning	**Unsupervised Learning**
Advantages	1. Uses Classes which are easy to understand	1. Detects unusual behaviour 2. Can discover new patterns

	2. Classes make it easy for pattern classification	3. Can be effectively used with publicly available unlabelled data (Bauder, Khoshgoftaar, & Seliya, 2017)
Disadvantages	1. Difficult to gather class labels 2. Costly to label input data 3. Cannot detect new types of frauds 4. Publicly available data has no class labels (Bauder, Khoshgoftaar, & Seliya, 2017)	1. Cannot detect duplicate claims 2. Instances exist where no interesting knowledge is gained 3. Includes uncertainty due to unknown links between the measured attributes (Bauder, Khoshgoftaar, & Seliya, 2017)

Hybrid Learning. Hybrid Learning basically refers to the combination of supervised and unsupervised learning to improve the fraud detection results (Bauder, Khoshgoftaar, & Seliya, 2017). This method combines the pros of the supervised and unsupervised learning methods. Hybrid is sometimes also referred as Semi-Supervised learning when some of the data used is labelled and some is unlabelled (Bauder, Khoshgoftaar, & Seliya, 2017). According to Bauder et al. (2017), hybrid learning can mitigate the risks associated with supervised and unsupervised learning when used on usually unlabelled, healthcare data. Hybrid learning also reduces the inherent risk associated with creating new labels for healthcare data (Bauder, Khoshgoftaar, & Seliya, 2017). Abdallah et al. (2017) describes the main aim of hybrid learning to train the model for both labelled and unlabelled data.

Use of Data mining algorithms in Health Insurance Fraud Detection

Yoo et al. (2011) described Classification, Clustering, and Association as the three most widely used data mining algorithms. Classification algorithm is a Supervised

(Predictive) Learning method whereas Clustering and Association are Unsupervised (Descriptive) Learning methods (Yoo, et al., 2011).

Liu & Vasarhelyi (2013) also provided a summary of fraud detection methods in the healthcare domain categorized on the basis of supervised, unsupervised and hybrid learning methods.

Table 2. Healthcare fraud detection methods (Liu & Vasarhelyi, 2013)

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

Joudaki et al. (2015) enlisted the application of various supervised, unsupervised and hybrid methods to the health insurance fraud detection domain.

Table 3. Data mining methods used in healthcare fraud detection domain (Joudaki, et al., 2015)

	Methods
Supervised Learning	• Decision Trees • Neural Networks • Support Vector Machine (SVM)
Unsupervised Learning	• Clustering • Outlier Detection • Association rules
Hybrid Learning	• Clustering and Rule Induction • Outlier Detection and Rule Extraction

Challenges in Health Insurance Fraud Detection

According to Sr. Solutions Architect, Ben Wright, at SAS, the major challenge faced by the analytics professionals relates to the quality of data available for processing (Wright, 2015). Data present maybe be accurate for local purpose, but it can be different for use across the organization. So, data integration and disambiguation pose as a challenge. This paper describes that many organizations have adopted data management software solutions to ensure accurate data in order of advanced analytics application. Another challenge is the timeliness and detail orientation of the data available for use. Data available to use is usually stored in data warehouses where it has been manipulated or summarized and, can be old as well. This paper describes that for efficient application

of data analytics, the operational data needs to be unaltered and provided on timely basis as close to its point of creation as possible (Wright, 2015).

Li et al. (2007) described the dynamic nature of health care data to be a challenge in efficient health insurance fraud detection system. Criminals are very adaptive and constantly keep adjusting their ways of committing fraud in order to elude the existing fraud detection techniques. There is a need for a 'self-learning' and 'self-evolving' fraud detection system in order to better adapt the varying patterns of fraudulent and legitimate claims (Li, Huang, Jin, & Shi, 2007).

(RGA, n.d.) discussed that in times when technology has improved the analytics capabilities for better fraud prevention and detection, it has also increased criminal's access to powerful data-based technologies which can lead to simplification of ways to scam health care programs. Fraudsters develop ways to work around the already existing strategies to thwart the fraud detection techniques. Also, fraudsters take the advantage of the reluctance of insurance companies to go ahead the with prosecution of fraud instances due to the high costs of litigation and the uncertainty involved in the outcomes.

Research Methodology

The purpose of this research project is to propose the features of a conceptual fraud detection approach in order to better detect the instances of fraud in the North American Health Insurance Industry. Furthermore, this research deliverable also included a side-by-side comparison and analysis of the major, existing fraud detection technologies predominantly being used in Canada.

The scope of the research revolved around using Intelligent Fraud Detection Technology to detect the instances of fraudulent claims in the health insurance sector.

Different approaches of data mining application solutions were studied related to the domain of health insurance fraud detection. The research only pertained to the use of fraud detection application solutions predominantly used in the North American region. The research deliverable was based on theoretical findings from the open-source resources during the research.

This research aimed at providing insights to the following research questions: (a) How can Fraud Detection technology be used more effectively in order to better detect the instances of fraudulent claims in healthcare insurance? (b) How can health insurance fraud detection application solutions be compared in order to determine the most effective solution? (c) What are the common features and capabilities of current healthcare insurance fraud detection application solutions?

The research study conducted was limited by the following factors: (a) The research is limited to the use of health insurance industry-specific applications. The research does not cover detection of fraud in other areas such as financial statement fraud, credit card fraud and/or telecommunication fraud; (b) The research does not consider the use of fraud detection application solutions outside the North American region such as the European or Asian markets; (c) The research is limited to the use and comparison of open-source fraud detection application solutions only; (d) The proposed research deliverable is theoretical, and no real-life implementation has been made to test for effectiveness.

As part of the research process, a literature review was used to review the existing literary works done in the field of healthcare insurance fraud detection to gain a deeper understanding of the underlying concepts. A side-by-side comparison of three existing

application solutions for healthcare insurance fraud detection was performed using a criterion that had been developed for the same. This comparison was based on the features, functionalities and capabilities of the application solutions as identified during the literature review process and using the information available in the public domain. Ultimately, the desired features of an application solution were determined for a more effective healthcare fraud detection conceptual approach.

Presentation of Results

Detecting fraudulent and abusive cases in healthcare is one of the most challenging problems for datamining studies (Kose, Gokturk, & Kilic, 2015). Healthcare Fraud has attracted many initiatives from the following three industries: healthcare, the data analytics, and the research communities, to work on fraud detection systems (Liu J. , et al.). There are a number of healthcare fraud detection vendor solutions present in the market today. But the vendor solutions chosen for this research were selected on the basis of a survey conducted by the *Market Research Future*, and an article by *Emerj.* The survey conducted by Market Future Research discusses the global healthcare fraud detection market segmentation based on a number of factors such as type, component, delivery model, application, region etc. (Market Research Future, 2019). The article, *AI for Health Insurance Fraud Detection,* discussed some of the artificial intelligence-based fraud detection vendor solutions available (Mejia, 2013). The following healthcare fraud detection application solutions were selected: SAS Viya: Detection and Investigation for Healthcare, H2O.ai: Driverless AI, and IBM Counter Fraud Management.

SAS Viya: Detection and Investigation for Healthcare

SAS Viya provides an end-to-end framework and workflow for detection, prevention and management of health care claims fraud. It includes components for fraud detection, alert management and case handling. (SAS Product Brief, n.d.). The solution’s fraud analytics engine uses various methods such as automated business rules, outlier analysis, predictive modeling, text mining, database searches, exception reporting, network link analysis, etc.to detect the instances of fraud that result in loss, and compromise payment integrity. Prioritized alerts are directed to investigators, auditors, provider payment

specialists and other business units, where the analysts can make use of the case management tools for more efficient investigation (SAS Product Brief, n.d.).

When a claim has been scored and prioritized, the analyst indulges in a deeper review of the claim to decide if the claim is fraudulent or not (SAS Product Brief, n.d.).

DentaQuest, CZ and Highmark Health are some of the clients that have used SAS for healthcare fraud detection (Mejia, 2013).

H2O.ai: Driverless AI

H2o.ai is a an open-source distributed in-memory artificial intelligence and machine learning based platform (Radhakrishnan, 2020). H2o.ai provides a machine learning platform which helps businesses and healthcare providers in creating their own AI based software. The Driverless AI platform has the ability of training itself and developing features (Mejia, 2013).

A customer can use the h2o.ai's platform to build a model based on a specific use case. In order for a user to build a machine learning model for healthcare fraud detection, the platform needs to be trained on a number of data points. The machine learning model is therefore exposed to a set of labeled health insurance claims. This trains the algorithm to be able to figure which of the claims can be fraudulent. After the deployment the client's company would be able to flag the fraudulent transactions as they appear in the system.

H2O.ai lists Change Healthcare, Armada Health, Kaiser Permanente and HCA to be some of the health industry clients (Pandey, 2020).

IBM Counter Fraud Management

IBM Counter Fraud Management (ICFM) is a single, integrated solution that provides enterprise counter-fraud measures (IBM Counter Fraud Management Documentation, n.d.). ICFM comprises of a management environment for the investigation, mitigation, and prosecution of fraud and financial crimes.

The Counter Fraud user interface includes built-in roles and processes which are required to triage, investigate, analyze, compile, and disposition the fraud investigations. Multiple user roles participate in this workflow (IBM Counter Fraud Management Documentation, n.d.). IBM Counter fraud management can be accessed through a web browser. Accessibility depends on the role assigned to the individual user.

The Medical Provider Fraud content pack CFAR file contains extensions, business objects, analytics, and sample data. The CFAR file is deployed to add the medical fraud detection features to the IBM Counter Fraud Management.

IBM lists the use of IBM analytics to spot suspicious activity and fraud among Medicaid patients and providers in North Carolina in the US.

The above-mentioned application solutions were compared with each other against some of the metrics that were selected based on the information available publicly and the relevancy to the research project. Some of the metrics used for the comparison are as follows: deployment methods, performance capabilities, security, reporting capabilities, risk assessment, user support and documentation etc. The following URL provides a detailed comparison of the three application solutions selected as a part of this research project:https://drive.google.com/file/d/1vIc_MzyIAbZHuSWd2fNuknCDQuLJuHuh/view?usp=sharing

The following are some of the desired features in an effective health care insurance fraud detection application solution based on the research conducted:

1. The application solution should be supported by the commonly used operating systems such as: Windows, Linux and Apple OS.
2. The application solution should consort to the latest accessibility features according to the W3C Web Content Accessibility Guidelines (WCAG). This makes the application solution accessible to people with a diverse range of hearing, movement, sight, and cognitive ability (W3C, n.d.).
3. The application solution should be able to process unstructured data.
4. The application solution should be able to aid in the investigation process by keeping track of the evidence in each fraud case.
5. The application solution should be integrated with a fraud case management solution.
6. The application solution should ensure the security of the information involved. Some of the common methods to enable data security are: Authentication, Authorization, Encryption and Web Security.
7. The application solution should be able to be tailored according to the needs and requirements of a business organization. Customization of the features according to use for specific use case should also be enabled.
8. The application solution should provide technical support in terms of documentation, knowledge base, online, over email and over phone.
9. The application solution should include social network analysis which can be useful in uncovering collusion rings.

10. The application solution should have the ability to analyse transactions on the basis of likelihood of fraud occurrence. The transactions should then be assigned a fraud risk score. Alerts should be generated for the investigator.
11. The application solution should include alert management which will help the investigator to prioritize the alerts.

Recommendations and Conclusions

Fraud is one of the major concerns for the health care insurance sector. The losses incurred due to fraud, like in any other industry, are very disruptive for the cost efficiency of the healthcare systems. This paper demonstrates the capabilities of intelligent fraud detection technology for detecting fraudulent claims in the health insurance domain.

The research aimed to list the features of an ideal health insurance fraud detection application solution. These features were derived from the comparative study of the selected vendor solutions. Some of the prominent features of an even more effective healthcare fraud detection vendor solution are integration with a fraud case management solution, ability to process unstructured data, and ability of the solution to be tailored according the business requirement.

The research describes how data mining is being used as an underlying concept in many of the fraud detection techniques that are being used in the field of health insurance fraud detection domain. The features and capabilities of the existing healthcare insurance fraud detection application solutions have been described as a part of this research. The main research deliverable, a side-by-side comparative study of three selected application solution, can be used as a basic criterion to compare one or more healthcare fraud detection vendor solutions. Also, the proposed features of a healthcare insurance fraud detection solution can be used to create a new detection framework in order to better detect the instances of healthcare fraud.

Although a number of intelligent software solutions exist today for the purpose of healthcare fraud detection, it has been found that more research is required into testing the effectiveness of the solutions. Whereas sectors such as credit card fraud or

telecommunication fraud have been well researched, a gap has been observed in the amount of research done in health insurance fraud area. This paper can pose as a base to future research since the comparative study done has been based only on the open-source material. The comparison of the selected software solutions can be explored further by practically testing the software solutions against a relevant dataset of healthcare data.

Intelligent Fraud Detection technologies can be used to facilitate more effective and efficient healthcare fraud detection solutions. Healthcare fraud detection solutions can therefore contribute to reduce the losses caused by the fraudulent health insurance claims.

Bibliography

(n.d.).

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud Detection System: A Survey. *Journal of Network and Computer Applications*, 90-113.

ACFE. (2018). *2018 Report to the Nations.* ACFE. Retrieved from acfe.com: https://www.acfe.com/uploadedFiles/ACFE_Website/Content/documents/cfe-employer-brochure.pdf

Ai, J. (2008). *Supervised and Unsupervised PRIDIT for Active Insurance Fraud Detection.* Austin: University of Texas.

Ai, J., Lieberthal, R. D., Smith, S. D., & Wojciechowski, R. L. (2018). Examining Predictive Modeling-based Approaches to Characterizing Health Care Fraud. *Society of Actuaries*.

Alberta BlueCross. (n.d.). *what are the consequences of commiting fraud?* Retrieved from Alberta BlueCross Fraud Prevention: https://ab.bluecross.ca/aboutus/fraud-report.php

AMA & CSMS. (2013). *the benefits of electronic claims submission - improve practice efficiencies.*

Bauder, R., Khoshgoftaar, T. M., & Seliya, N. (2017). A Survey on the state of healthcare upcoding fraud analysis and detection. *Health Services & Outcomes Research Methodology*, 31-55.

BlueCross, A. (2016). *Fraud Prevention.* Retrieved from Alberta Blue Cross: https://www.ab.bluecross.ca/pdfs/81984-fraud-member-letter.pdf

Busch, R. S. (2007). *Healthcare Fraud: Auditing and Detection Guide.* John Wiley & Sons, Inc.

D'Agostino, G., & Woodward, D. A. (2010). Diagnosing Our Health Records in the Digital World: Towards a Legal Governance Model for the Electronic Health Record in Canada. *Intellectual Property Journal, Toronto*, 127-154.

Evbayiro, H. O. (2011). A historical analysis of federal policies on health care fraud. *Available from ProQuest Dissertations & Theses Global*.

FBI. (2006). *Financial Crimes Report to the Public.*

Harrison, J. P., & Lee, A. (2006). The role of E-health in the changing health care environment. *Nursing Economics*. Retrieved from CNE OBJECTIVES AND EVALUATION FORM: https://search-proquest-com.ezproxy.aec.talonline.ca/docview/236937602/fulltextPDF/34403ACC284842C2PQ/1?accountid=10243

HHS. (2018). *Health care fraud and abuse control report.* US Department of Health and Human Services.

Hyman, D. A. (2002). HIPAA and Health Care Fraud: An empirical perspective. *Cato Journal*, 151-178.

IBM Counter Fraud Management Documentation. (n.d.). *IBM Counter Fraud Solution Overview*. Retrieved from ibm.com: https://www.ibm.com/support/knowledgecenter/SSXJFX_2.0.0.3/productoverview.html

IBM Counter Fraud Management Documentation. (n.d.). *Using the IBM Counter Fraud Management interface* . Retrieved from ibm.com:

https://www.ibm.com/support/knowledgecenter/SSXJFX_2.0.0.3/t_gettingstarted.html

Jain, A., & Shinde, S. (2019). A Comprehensive Study of Data Mining-based Financial Fraud Detection Research. *2019 IEEE 5th International Conference for Convergence in Technology(12CT)* . Pune.

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Using Data Mining to Detect Health Care Fraud and Abuse: A Review of Literature. *Global Journal of Health Science*, 194-202.

Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Elsevier*.

Kou, Y., Lu, C.-T., Sirwongwattana, S., & Huang, Y.-P. (2004). Survery of fraud detection techniques. *IEEE International Conference on Networking, Sensing and Control*.

Li, J., Huang, K.-Y., Jin, J., & Shi, J. (2007). A survey on statistical methods for health care fraud detection . *Springer Science + Business Media*.

Liu, J., Bier, E. W., Guerra-Gomez, J. A., Honda, T., & al., e. (2016). Graph Analysis for Detecting Fraud, Waste, and Abuse in Health-Care Data. *AI Magazine*, 33-46.

Liu, J., Bier, E., Wilson, A., Honda, T., Kumar, S., Gilpin, L., . . . Davis, D. (n.d.). Graph Analysis for Detecting Fraud, Waste and Abuse in Healthcare Data. *Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence.* Palo Alto.

Liu, Q., & Vasarhelyi, M. (2013). Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information. *29th WORLD CONTINUOUS*

AUDITING AND REPORTING SYMPOSIUM (29WCARS), BRISBANE, AUSTRALIA.

Market Research Future. (2019, September). *Global Healthcare Fraud Detection Market*. Retrieved from Marketresearchfuture.com: https://www.marketresearchfuture.com/reports/healthcare-fraud-detection-market-5670

Matloob, I., Khan, S., Rahman, H. u., & Hussain, F. (2020). Medical Health Benefit Management System for Real-Time Notification of Fraud using Historical Medical Records. *Applied Sciences*.

Mejia, N. (2013, December 13). *AI for Health Insurance Fraud Detection - Current Applications*. Retrieved from Emerj: https://emerj.com/ai-sector-overviews/ai-for-health-insurance-fraud-detection-current-applications/

NHCAA. (2018). *The challenge of health care fraud.* Retrieved from nhcaa.org: https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud/

Office of Public Health. (2020, September 30). *US Department of Justice.* Retrieved from https://www.justice.gov/opa/pr/national-health-care-fraud-and-opioid-takedown-results-charges-against-345-defendants

Pandey, P. (2020, March 23). *How H2O.ai is reinventing healthcare with AI*. Retrieved from h2o.ai: https://www.h2o.ai/blog/how-h2o-ai-is-reinventing-healthcare-with-ai/

Piper, C. (2013, January). *10 popular health care provider fraud schemes*. Retrieved from ACFE: https://www.acfe.com/article.aspx?id=4294976280

Radhakrishnan, R. (2020, April 7). *Unleashing the power of AI with optimized architectures for H20.ai*. Retrieved from proquest.com: https://www.proquest.com/docview/2387109798/fulltext/B8509BA89CE9443CPQ/1?accountid=10243

Rawte, V., & Anuradha, G. (2015). Fraud detection in health insurance using data mining techniques. *2015 International Conference on Communication, Information & Computer Technology(ICCICT)*. Mumbai.

RGA. (n.d.). *Eye on Fraud:Current State*. Retrieved from rga: https://www.rgare.com/docs/default-source/brochure/fraud_whitepaper_v6.pdf?sfvrsn=9f642f0f_0

SAS Product Brief. (n.d.). *SAS Detection and Investigation for Healthcare*. Retrieved from sas.com: https://www.sas.com/content/dam/SAS/en_us/doc/productbrief/sas-detection-investigation-health-care-104106.pdf

Segal, S. Y. (2016). Accounting frauds - review of advanced technologies to detect and prevent frauds 1. *Economics and Business Review*, 45-64.

Seth, V. (n.d.). *Healthcare Privacy Legislation*. Retrieved from colleage: https://www.colleaga.org/article/healthcare-privacy-legislation-canada

Stowell, N. F., Schmidt, M., & Wadlinger, N. (2018). Healthcare Fraud under the microscope; improving its prevention. *Journal of Financial Crime*, 1039-1061.

Thornton, D., Brinkhuis, M., Amrit, C., & Aly, R. (2015). Categorizing and Describing the Types of Fraud in Healthcare. *Conference on ENTERprise Information System.*

Thornton, D., Mueller, R. M., Schoutsen, P., & Van Hillegersberg, J. (2013). Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection. *Procedia Technology* , (pp. 1252-1264).

Van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., & Van Hillegersber, J. (2016). Outlier Detection in Healthcare fraud: A case study in the Medicaid dental domain. *International Journal of Accounting Information Systems*, 18-31.

W3C. (n.d.). *W3C Accessibility*. Retrieved from w3.org: https://www.w3.org/standards/webdesign/accessibility

Welch, S. T., Holmes, S. A., Strawser, R. H., & Evans, S. M. (1998). Fraud in the health insurance industry. *Jounral of the American Society of CLU & ChFC*, 70-76.

West, J., & Bhattacharya, M. (2016). Intelligent Financial Fraud Detection: A comprehensive. *Computers & Security*, 47-66.

Wright, B. (2015, November). *Health Care Payment Integrity through Advanced Analytics.* Retrieved from NHCAA: https://www.nhcaa.org/media/93915/iia-sas-payment-integrity-white-papers.pdf

Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.-F., & Hua, L. (2011). Data Mining in Healthcare and Biomedicine: A Survey of the Literature. *Springer Science*, 2431-2448.