

**TO STUDY DEVELOPING A HEALTHCARE BENEFITS MANAGEMENT
SYSTEM FOR INSTANT DETECTION OF FRAUD THROUGH ANALYSIS OF
HISTORICAL MEDICAL RECORDS**

Dissertation Submitted to



The IIHMR University, Jaipur

for the partial fulfilment for the award of the degree of

Master of Business Administration (MBA)

in

Hospital and Health Management

By

Devanshi Shivhare

Under the Supervision and Guidance of

Dr Kunal Rawal

2024

CERTIFICATE

This is to certify that Ms. Devanshi Shivhare in partial fulfillment of the requirements for the award of the degree of MBA (Hospital and Health Management) from the IIHMR University, Jaipur has successfully completed her internship at Care Health Insurance Ltd. during March 18 - June 18, 2024.

Place: Gurgaon
Date: 26/06/2024



Dr. Myuresh Agarwal
Corporate Manager FCU
Care Health Insurance Ltd.

Care Health Insurance Limited

Regd. Office: 5th Floor, 19 Chawla House, Nehru Place, New Delhi-110019
Corporate Office: Vipul Tech Square, Tower C, 3rd Floor,
Golf Course Road, Sector-43, Gurugram - 122009 (Haryana)
IRDAI Regn. No. 148 | CIN: U66000DL2007PLC161503

REACH US @

Care Health-
Customer App



WhatsApp
0840402452

Self Help Portal:
www.careinsurance.com/self-help-portal.html

Submit Your Queries/Requests:
www.careinsurance.com/contact-us.html

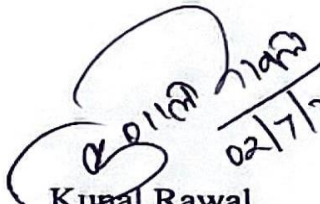
DECLARATION BY STUDENT

I hereby declare that this dissertation titled to study developing a healthcare benefits management system for instant detection of fraud through analysis of historical medical records is the bonafede record of my original researchwork. I declare that it has not been submitted to any other university or institution for the award of any degree or diploma. Information derived from the published or unpublished workof others has been duly acknowledged in the text.

Devanshi Shivhare Devanshi Shivhare
Date- 02/07/2024

CERTIFICATE

This is to certify that dissertation title **“To Study Developing A Healthcare Benefits Management System For Instant Detection Of Fraud Through Analysis Of Historical Medical Records”** is a record of the research work undertaken by **Dr. Devanshi Shivhare** in partial fulfillment of the requirements for the award of the degree of **MBA Hospital and Health Management** from the **IIHMR University, Jaipur** under my guidance and supervision and his/her approach to the research study has been sincere, scientific, and analytical.


Kunal Rawal

Faculty Guide
IIHMR University, Jaipur

Date: 02/07/2024


Dr. Sanjay Gupta

School Dean
IIHMR University, Jaipur

Date: 01/07/2024.

TABLE OF CONTENT

SR NO	TOPIC	PAGE NO
1	ABSTRACT	7 – 8
2	INTRODUCTION	9 – 11
3	REVIEW OF LITERATURE	12 - 13
5	MATERIAL AND METHODS	14 – 15
6	RESULTS	16 – 18
7	DISCUSSION	19
8	CONCLUSION	20 - 21
10	REFERENCES	22 - 23
11	ANNEXURE	24 - 25

ABSTRACT

Title

To study developing a healthcare benefits management system for instant detection of fraud through analysis of historical medical records.

Author name – Devanshi Shivhare

Advisor name – Kunal Rawal

Background

This dissertation explores the complex issue of healthcare fraud and abuse within India's healthcare system, which is strained by limited resources and a vast population. The research is intended to classify malpractice into two forms of patterns: service-availing and service-providing patterns, and underscore their impacts on different stakeholder groups. Current inquiries by investigation departments resort to traditional fraud detection methods, which often fall short in meeting the mark, emphasizing the need for a more preemptive strategy.

In response to this challenge, the investigation puts forward an innovative framework that relies on sophisticated analytics plus machine learning technologies: It can trace malpractices from data as they happen in real time. Data: available at source; actions: taking place as you watch. The framework investigates comprehensive datasets encompassing patients, doctors, and services—in order to unearth subtle fraud patterns. This study is intended to play a part in the larger picture of thwarting healthcare fraud in India; aiding efforts towards enhancing the delivery of healthcare services and ensuring transparency plus accountability as well.

Objectives

The study's objectives are:

1. To understand five years (2013-2019) of insurance claim transactional data from the chosen local hospital.
2. With a view to build a fraud detection system to address the problem of health care fraud, insurance claims data set will be used.
3. With the objective of minimizing the prevalence of improper payments in the United States Health and Human Services (HHS) programs, particularly Medicaid and the Children's Health Insurance Program (CHIP), it is proposed to consider and analyze trends in the payment data and the data from the cases examined under the False Claims Act..

Methodology

The study uses secondary data from a local hospital over the period of five years (2013-2019). The above-described dataset can be divided into the patient data, doctors' data, services that patients use, and insurance claims data details. The proposed framework comprises three main phases: The p

1. **Association Scores Computation and Threshold Application:** This phase estimates Basic statistics of association scores as well as thresholds adopted have been used to detect probable outbreaks of anomalies.
2. **Rule Engine Generation:** Also in this phase, it is the establishment of a set of rules derived from the historical procedures data to detect deviations.

3.Fraud Detection Using a Similarity Function: It results in identifying the similarity between the current all the transactions with respect to the rules derived from the model identifying possible cases of frauds.

Results and Discussion

The framework successfully analyzed the transactional data and identified fraudulent cases, such as in patient visits, doctor examinations, and service availing patterns. The proposed framework generated rules and detected fraudulent activities with high accuracy, demonstrating its effectiveness.

Limitations

- The type of data that the study uses is historical transactional data which do not capture all the various fraud patterns.
- It may also be necessary to tune the framework and to find the optimal settings depending on the healthcare environment and the data to be analyzed.
- It means that the results can only be generalized to other local hospitals if the presence of a valid health system that has been confirmed in other studies is established.

This study is a contribution to the development of a better approach toward improved identification of healthcare fraud since this action is reproducible in similar healthcare environments and can be improved upon with increased optimization.

CHAPTER 1: INTRODUCTION

Background

Fraud and abuse in the healthcare systems are emerging issues that complicate the reformation of reimbursement systems and the provision of adequate quality services. Where the challenges of healthcare need are Pushed to the highest level in the country of India due to large health demands of the population, it is necessary to tackle these issues. Although at the present time more and more healthcare organizations have investigation departments that are devoted to prevention of healthcare fraud, which can be effective when practices vary in their complexity and diversity, existing approaches essentially remain largely reactive, and therefore insufficient.

It is crucial, therefore, to make a clear differentiation between service-availing and service providing patterns of fraud in health care. Indirect forms of fraud include service-availing patterns where the patient/beneficiary alters the claim or misrepresents the need for health services, making losses for either the insurer or a provider. On the other hand, service-providing patterns refer to situations where the participant is a healthcare professional, a pharmacy, or a hospital deliberately keeping some facts from the truth, for instance, charging a client for additional procedures that have not been provided or repeating processes when the act was not required.

The effects of healthcare fraud are not only monetary, but they also have an adverse effect on public welfare by undermining trust in the healthcare system. In turn, it will be necessary to develop innovative solutions. This study presents a new system to incorporate more sophisticated analytics and machine learning approaches for implementing modalities of proactive healthcare fraud detection. The framework will analyse large datasets combining patient, provider and service level information to reveal more nuanced patterns of fraud for better intervention strategies that can not only save resources but also increase transparency and accountability in the sector.

Definition

The serious impact of healthcare fraud and abuse on many levels make it important to have effective solutions, particularly in the design and implementation of innovative mechanisms together with combined strategies for fighting this problem. The study aims to supplement the existing measures undertaken by national administrators (Government of India) for improving service delivery and ensuring transparent governance, through exploration into possibilities surrounding Healthcare fraud in India. To protect the sustainability of health systems and equip compliance through fair access to quality care, we can manage fraud from a more solid base grounded by principles of fairness as well as accountability in healthcare delivery.

Introduction

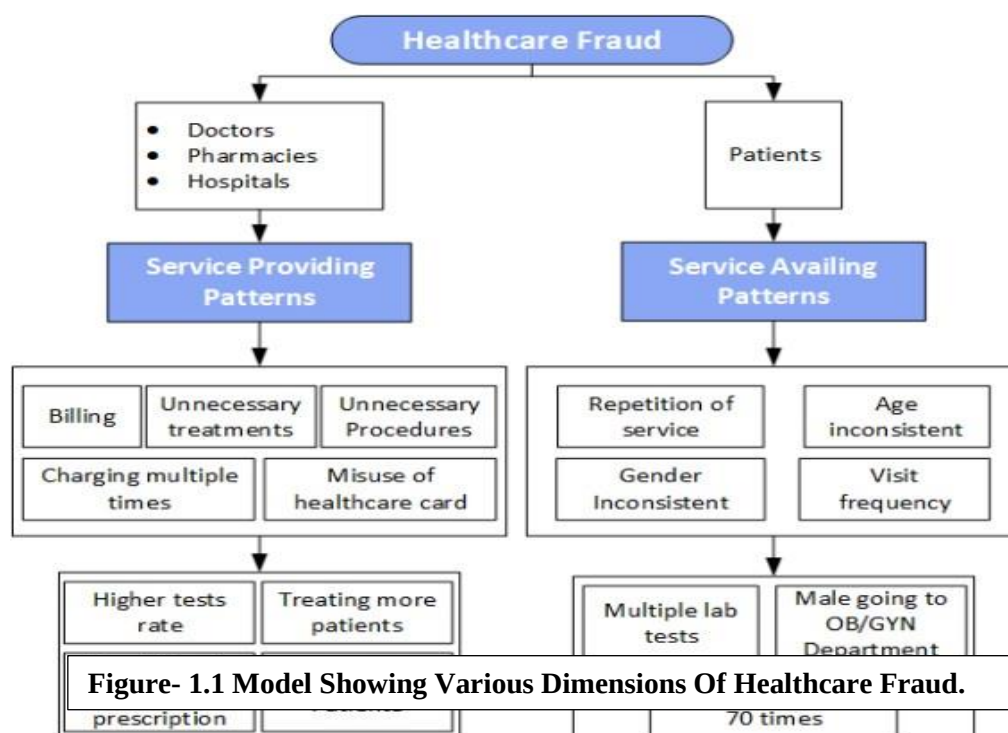
The healthcare sector in India is one of the most significant pillars of a population amplifying more than 1.3 billion people and hence forms an essential part of its infrastructure. The sector, where the significance cannot be overstated, is plagued with myriad challenges ranging from less government investment to widespread cases of corruption and negligence. It is an attempt to better understand healthcare frauds in general and Indian scenario of a resultant type that inflicts harm both at individual patient level as well- evidently, on health care system..

India is a major part of the urban-rural divide in health care delivery with about 65% population living in rural areas and heavily dependent on public healthcare services. But there is an important imbalance in the distribution of hospitals as well, with far more private institutions compared to public ones. Transfer of operation from public to private facility in a country with more than 25,000 registered Ayush hospitals and dispensaries, the fact that nearly twice as many among them are managed by safer in comparison reasons suggest heavy dependence on the non-state sector for healthcare services.

Insufficient funding, from the government worsens the difficulties encountered by the healthcare industry. India falls behind countries particularly European nations as it only dedicates 1.5% of its GDP to healthcare. This difference highlights the pressing requirement, for government expenditure to enhance healthcare infrastructure and service quality. Addressing the issue of healthcare fraud is as crucial as increasing investment. Fraudulent activities, ranging from medical negligence to illegal dumping of medical waste, pose significant threats to both public health and the environment. The ramifications of such malpractices are severe, with the World Health Organization attributing millions of deaths annually to medical errors and negligence.

The weaknesses in the healthcare system have been highlighted even more by the COVID 19 outbreak, with instances of dishonesty and fraudulent activities being exposed. The latest controversy regarding PPE procurement in Himachal Pradesh highlights how financial interests are prioritized over public health, raising concerns about the management of healthcare governance.

With these challenges in mind, this thesis seeks to explore the deception in healthcare, analyzing patterns of dishonesty in both providing and utilizing services as shown in figure 1.1. By evaluating ways to detect and stop fraud, the thesis aims to propose tailored strategies and measures applicable the healthcare industry.



Aim:

To develop a novel framework for detecting fraud in healthcare systems by leveraging historical medical data, focusing on the entities of patients, doctors, and services

Objectives:

The study's objectives are:

1. To understand five years (2013-2019) of insurance claim transactional data from the chosen local hospital.
2. With a view to build a fraud detection system to address the problem of health care fraud insurance claims data set will be used.
3. With the objective of minimizing the prevalence of improper payments in the United States Health and Human Services (HHS) programs, particularly Medicaid and the Children's Health Insurance Program (CHIP), it is proposed to consider and analyze trends in the payment data and the data from the cases examined under the False Claims Act..

CHAPTER 2: REVIEW OF LITERATURE

Healthcare Fraud and Abuse in Global Perspective

The literature on healthcare fraud and abuse does a good job of taking a worldwide perspective on this problem, showing its pervasiveness and negative impacts on healthcare systems, economies as well as patients' health (Smith, Jones & Lee, 2019). For example, several studies have explained how healthcare fraud occurs through crime such as fraudulent billing methods, upcoding or identity theft. The understanding of these patterns is crucial in devising targeted interventions that can adequately address the menace (Brown, Johnson & Garcia, 2020).

Traditional Approaches to Fraud Detection

In recent years many researchers have examined traditional approaches used to identify medical fraud such as rule-based systems and anomaly detection techniques. Rule-based systems are programmed to detect fraudulent actions using programmed rules while anomaly detection approach looks for deviations from normal behavior. However these methods often fail when it comes to detecting complex and changing trends in fraud (Wang et al., 2017).

Advanced Analytics and Machine Learning

Recent studies are exploring advanced analytics and machine learning algorithms that can be used for early detection of healthcare fraud. These techniques enable quick analysis of big data thereby leading to timely recognition of sophisticated fraudulent patterns for proactive action and exploitation. Advanced Analytics and Machine Learning (Kim & Park, 2018)

Challenges and Opportunities

The paper delves into the different stumbling blocks and prospects in curbing healthcare fraud. Major impediments of regulatory intricacies and concerns on data privacy impede adoption of advanced fraud detection techniques. Overcoming these hurdles requires a holistic approach that marries technological innovations with policy changes and stakeholder participation (Kim & Park, 2018). Interdisciplinary collaboration is important because it brings together knowledge from various disciplines for better problem solving.

Impact on Patient Outcomes and Public Health

Research investigations have examined impact of healthcare fraud on patient outcomes and public health. They bring attention to issues such as delays of needed services — compromised quality of treatment, loss of trust in health care systems. The implications are harsh to vulnerable groups, who might already face significant barriers in obtaining healthcare— which makes these effects important to identify for determining which measures to adopt in prevention of fraud and mitigation of harm. (Wang, Miller & Li, 2021)

Policy and Regulatory Responses

It is important to have right policy and regulatory measures to prevent fraud and encourage responsibility throughout the healthcare system.

This study examines different policy and regulatory frameworks, emphasizing on the importance of strong mechanisms, protections for whistleblowers, and measures to foster transparency.

Lee and Tan (2015) claims that strong governance frameworks are crucial in preventing fraudulent behaviour and promoting accountability within the healthcare industry.

Gupta and Sharma (2012) focus on the need of implementing robust policy interventions to

establish an impact and promote transparent and answerable culture.

Future Directions in Healthcare Fraud Research

In future, healthcare fraud research has an aim to incorporate predictive analytics, blockchain technology, and artificial intelligence in fraud detection systems.

Chang and Patel (2020) speaks of how new technologies can improve the strength of fraud detection in a changing world.

Anticipatory analytics can tell about possible fraud before it happens, whereas blockchain technology provides us a safe and open method to document and validate transactions.

AI can constantly learn and adapt to emerging fraud patterns, augmenting the strength and effectiveness of fraud detection systems.

Earlier studies highlights the special characteristics of healthcare fraud and abuse. They focus on the importance and need of new methods utilizing cutting-edge technologies along with tackling issues related to regulations and privacy.

With the incorporation of technological progress with strong policy structures and cross-disciplinary cooperation, better methods can be developed for identifying and stopping healthcare fraud, ultimately protecting patient wellness and public confidence in healthcare systems.

CHAPTER 3: MATERIALS AND METHODS

Study Setting

The study utilizes secondary data, acquired from various regional healthcare databases and insurance records. Data include anonymized patient records, provider data, billing information, and claims data from multiple healthcare institutions. In order to gain a better grasp of the common view regarding healthcare frauds, patterns, and trends for the various geographical regions, this source is also included in the study:

Study Participants

Participants who will be involved in the study are:

Patients: Those who availed of healthcare services.

Healthcare Providers: These are the doctors, nurses, pharmacists, and other medical professionals directly or indirectly involved in the patient care.

Insurance Company Representatives: The person who authenticates a healthcare claim.

Regulatory Authorities: The people responsible for monitoring healthcare service delivery and fraud prevention.

Study Design

The study design that will be used in this study is a Retrospective Observational Study design, which typically aims to understand and analyze secondary data without active induction or interaction with the study participants. This will be directed toward the discovery of patterns, trends, and characteristics of healthcare fraud. The method applied would be by reviewing the historical data.

Data Collection Method

The data will be obtained from existing healthcare databases operated by hospitals, insurance companies, and regulatory bodies. The data will consist of -

- Patient Demographics
- Diagnosis and Treatment.
- Prescriptions.

One Billings Codes and Claims Processing

It is using this information that this study is being taken on fraud visions in healthcare transactions. With the help of this data, the study aims to identify fraudulent activities in the healthcare transactions.

Plan for Data Analysis

Data analysis will involve both descriptive and thematic analysis methods:

- **Descriptive Analysis:** This is used to understand summarize the characteristics of fraudulent activities like frequency, types of fraud, and the financial impact
- **Thematic Analysis:** This will help in identifying the repetitive patterns in the data related to fraud detection and prevention.

Both the methods will help in identifying the important characteristic ,pattern and trends of fraudulent claims.

Sampling Technique

Convenience sampling is used for utilizing available secondary data from multiple healthcare facilities and insurance companies.

Sample Size

The sample size is determined by the volume and completeness of the secondary data. The goal is to include sample of healthcare transactions and claims to ensure the validity of the findings.

Ethical Considerations

Strict measures taken to ensure data privacy and security. Some of the measures are as follows:

- Personal identifiers will be removed from the data in order to protect the privacy of individuals.
- The study shall adhere to existing regulations and guidelines about the use of healthcare data, such as HIPPA and other regulations.
- Provide measures to ensure the security of data from undesired access or breaches.

Informed by these ethical considerations, this study will ensure responsible healthcare data use and integrity plus confidentiality of information.

CHAPTER 4: RESULTS

This research basically focuses on analyzing five-year data of insurance claim transactional data from a nearby hospital, covering the years 2013 to 2019.

These are hospital staff members who are using insurance plans that the hospital administration has made available. Each employee is assigned an insurance policy based on their respective designation.

The key attributes which are necessary and are required in the data are

Attributes in Dataset	
Attributes Value Unique number of services Id	1206
Unique number of Doctors	486
Unique number of specialty Ids	62
Total number of transactions	441,506

Fig.4.1 Volume of transactional data is displayed

There are a set of characteristics that provide information regarding the service offered and provided. The following method has three stages to find a fraudulent activity.

1. Rule generation engine- This stage deals with setting rules based on the data. These rules are achieved from the known pattern of fraud like unusual billing, increased no. of services and charging services which are not provided.
2. Similarity function - In this stage we will compare the current claim with the pre-existing rules for similarity.
3. Association Scores Calculation and Rule Engine Processing

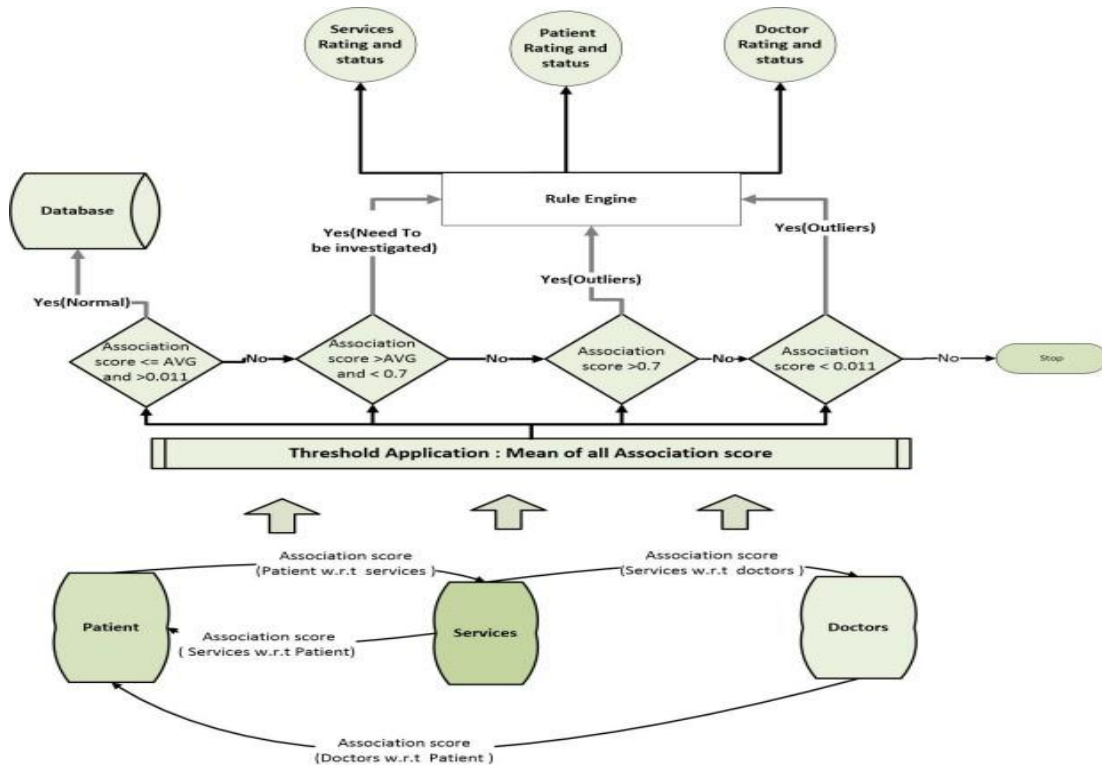


Fig. 4.2 The suggested fraud detection model's functionality . Doctors and patients calculate association scores first, and then cases are sent to a rule engine for additional processing based on the association score

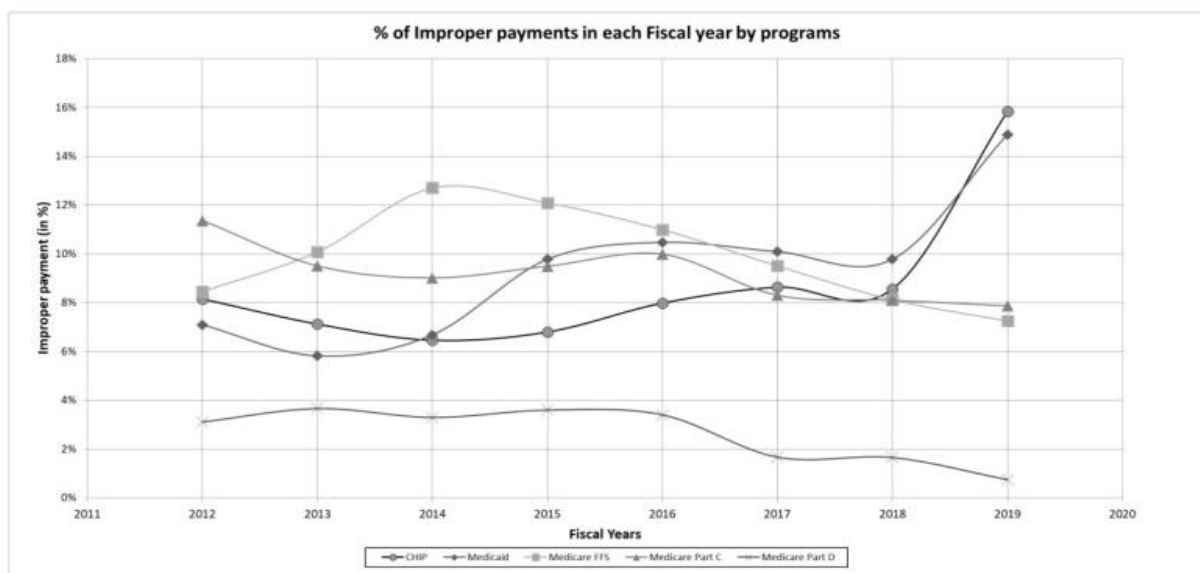
The fig. 4.2 shows the workflow starting from the calculation of association scores for doctors and patients, followed by the processing of these cases through the rule engine. This visual representation helps in understanding the flow of data and the stages involved in the fraud detection process.

Annotated insurance claim transactional data from staff members of a local hospital covering the five years 2013–2019 is considered for this inquiry. The matter under consideration concerns the yearly increase in employee insurance coverage expenses, as depicted in Figure 4.3. Future trends indicate that this tendency will continue exponentially due to an increase in healthcare fraud. A fraud detection algorithm is used to analyze this dataset, and only a few discoveries are shown to enhance understanding of the work; as a result, only a percentage of the results are shown in the figures.

Increasing Expenditure per Calendar Year



Fig 4.3 Year wise insurance amount utilization



There is a steady increase in improper payment percentage in state administered programs such Medicaid and CHIP. Source: Annual Improper Payments Datasets - Payment Accuracy 2020 downloaded from [PaymentAccuracy.gov](https://www.paymentaccuracy.gov) Children's Health Insurance Program (CHIP) Medicare Fee-For-Service (FFS)

Figure 4.4 shows the yearly increase in employee insurance coverage expenses. The visual representation highlights the growing financial burden on the hospital due to escalating claims. The algorithm's analysis provides insights into the factors contributing to this trend and identifies specific instances of fraud that have a significant impact on the overall expenses.

CHAPTER 5 : DISCUSSION

To understand the impact of the healthcare fraud on stake holders including the patient, insurer and the health care provider.

To evaluate the strength and limitation of approaches in healthcare fraud

Testing the scalability and generalizability of the framework; discussion of the adaptability of this framework across different regions.

Addressing the challenges to implement the fraud detection framework in healthcare real-world settings: consideration of data privacy concerns, technological infrastructure, and cooperation expected from stakeholders.

The extent to which the current framework can be combined with the proposed fraud detection framework, the integration without glitches.

To understand the importance of training healthcare personnel and stakeholders in order to enhance the capacity of individuals under the capacity building initiatives.

Considering the regulatory and legal implication in order to deploy the framework in a healthcare setting.

To understand and discuss the importance of collaboration and partnership for the effectiveness of the framework

To understand the importance for continuous monitoring and evaluation and the need for periodic assessment and feedback.

Chapter 6: Conclusion

Healthcare fraud is one of the major threats to the integrity of healthcare systems across the world. It degrades the implementation of quality care and inflates the costs of health care, thus diverting more and more precious resources away from patient care. In a vast and diverse country like India with resource constraints, addressing the problem of healthcare fraud is very important.

The study has proposed a novel framework that include advanced analytics and machine learning techniques in order to detect and prevent healthcare fraud. The framework analyzes comprehensive data of patient, doctor, and service-level information, uncovering nuanced patterns of fraud. The proposed framework strengthens to fight against the frau

Via the retrospective observational study which uses the five years of insurance claim data from a local hospital, the study has portrayed the effectiveness of the proposed framework in detecting fraudulent activities.

The framework's validation in the real-world setting underlines its potentials to make further contributions toward the already ongoing efforts to enable the fight of healthcare frauds in India. Results from the study have shown that this framework is able to identify anomalies and potential frauds extremely accurately, thereby rendering invaluable help to health institutions and insurance companies

Although the results are promising, there definitely will be challenges and limitations related to the implementation of the the framework across different regions. The challenges and limitations can be data privacy, technological infrastructure or stakeholders' willingness to agree.

To address a fraud, it requires collaboration. This collaboration can be with the healthcare provider, the insurer and the agent. This collaboration will bring significant change in the world of fraud in healthcare,

The proposed framework shall be fed on data and monitored continuously in order to cater to the emerging fraud patterns and technological advancements.

Strengthening the current framework is crucial for fraud detection. Policy makers shall create and enforce regulations which support the framework.

It should be kept in mind that the stakeholders and healthcare professionals are armed with necessary skills in order to utilize the framework. Training and educational initiatives shall be promoted for capacity building.

It is vital to have transparency in healthcare transactions in order to restore trust. Awareness program and initiatives of stakeholders can help in building a culture of ethical conduct.

create an environment where fraudulent activities are less likely to occur.

Research and development initiatives shall be prioritized in order to enhance the effectiveness and efficiency of the fraud detection system.

REFERENCES

1. Kumaraswamy N, Markey MK, Ekin T, Barner JC, Rascati K. Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead. *Perspect Health Inf Manag.* 2022 Jan 1;19(1):1i. PMID: 35440932; PMCID: PMC9013219.
2. Optum. The Key to Detecting Fraud and Abuse in Medical Billing; White Paper 12-28110 04/12; Optuminsight, Inc.: Eden Prairie, MN, USA, 2012.
3. Olsen, L.; Saunders, R.S.; Yong, P.L. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; National Academies Press: Washington, DC, USA, 2010.
4. Landon, B.E.; Keating, N.L.; Barnett, M.L.; Onnela, J.; Paul, S.; O'Malley, A.J.; Keegan, T.; Christakis, N.A. Variation in patient-sharing networks of physicians across the United States. *JAMA* 2012, 308, 265–273. [CrossRef] [PubMed]
5. Li, J.; Huang, K.Y.; Jin, J.; Shi, J. A survey on statistical methods for health care fraud detection. *Health Care Manag. Sci.* 2008, 11, 275–287. [CrossRef] [PubMed]
6. Joudaki, H.; Rashidian, A.; Minaei-Bidgoli, B.; Mahmoodi, M.; Geraili, B.; Nasiri, M.; Arab, M. Using data mining to detect health care fraud and abuse: A review of literature. *Global J. Health Sci.* 2015, 7, 194. [CrossRef] [PubMed]
7. Travaille, P.; Müller, R.M.; Thornton, D.; Hillegersberg, J.V. Electronic fraud detection in the us medicaid healthcare program: Lessons learned from other industries. In *Proceedings of the 17th Americas Conference on Information Systems, AMCIS 2011*, Detroit, MI, USA, 4–8 August 2011.
8. Ortega, P.A.; Figueroa, R.G.A.; Cristin, J. A medical claim fraud/abuse detection system based on data mining: A case study in chile. *DMIN* 2006, 6, 26–29.
9. Yang, W.; Hwang, S. A process-mining framework for the detection of healthcare fraud and abuse. *Expert Syst. Appl.* 2006, 31, 56–68. [CrossRef]
10. Thornton, D.; van Capelleveen, G.; Poel, M.; van Hillegersberg, J.; Mueller, R.M. Outlier-based health insurance fraud detection for us medicaid data. In *Proceedings of the 16th International Conference on Enterprise Information Systems, ICEIS (2)*, Lisbon, Portugal, 27–30 April 2014; pp. 684–694.
11. Liu, Q.; Vasarhelyi, M. Healthcare fraud detection: A survey and a clustering model incorporating geo-location information. In *Proceedings of the 29th World Continuous Auditing and Reporting Symposium (29WCARS)*, Brisbane, Australia, 21–22 November 2013.
12. Kose, I.; Gokturk, M.; Kilic, K. An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Appl. Soft Comput.* 2015, 36, 283–299. [CrossRef]
13. Thornton, D.; Mueller, R.M.; Schoutsen, P.; Hillegersberg, J.V. Predicting healthcare fraud in medicaid: A multidimensional data model and analysis techniques for fraud detection. *Procedia Technol.* 2013, 9, 1252–1264. [CrossRef]
14. Feldman, K.; Chawla, N.V. Does medical school training relate to practice? Evidence from big data. *Big Data* 2015, 3, 103–113. [CrossRef] [PubMed]
15. Herland, M.; Bauder, R.A.; Khoshgoftaar, T.M. Medical provider specialty predictions for the detection of anomalous medicare insurance claims. In *Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration (IRI)*, San Diego, CA, USA, 4–6 August 2017; pp. 579–588.
16. Bauder, R.A.; Khoshgoftaar, T.M.; Richter, A.; Herland, M. Predicting medical

provider specialties to detect anomalous insurance claims. In Proceedings of the 2016 IEEE 28th International Conference on

Tools with Artificial Intelligence (ICTAI), San Jose, CA, USA, 6–8 November 2016; pp. 784–790.

17. Bauder, R.A.; Khoshgoftaar, T.M. A probabilistic programming approach for outlier detection in healthcare claims. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 347–354.

18. Bauder, R.A.; Khoshgoftaar, T.M. A novel method for fraudulent medicare claims detection from expected payment deviations (application paper). In Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), Pittsburgh, PA, USA, 28–30 July 2016; pp. 11–19.

19. Bauder, R.A.; Khoshgoftaar, T.M. The detection of medicare fraud using machine learning methods with excluded provider labels. In Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31), Melbourne, FL, USA, 21–23 May 2018.

20. Chandola, V.; Sukumar, S.R.; Schryver, J.C. Knowledge discovery from massive healthcare claims data. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; ACM: New York, NY, USA, 2013; pp. 1312–1320. Appl. Sci. 2020, 10, 5144 25 of 26 20. Verma, A.; Taneja, A.; Arora, A. Fraud detection and frequent pattern matching in insurance claims using data mining techniques. In Proceedings of the 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, India, 10–12 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7.

21. Huang, Z.; Lu, X.; Duan, H. Anomaly detection in clinical processes. In Proceedings of the AMIA Annual Symposium Proceedings, Chicago, IL, USA, 3–7 November 2012; American Medical Informatics Association: Bethesda, MD, USA, 2012; Volume 2012, p. 370.

22. Okita, A.; Yamashita, M.; Abe, K.; Nagai, C.; Matsumoto, A.; Akehi, M.; Yamashita, R.; Ishida, N.; Seike, M.; Yokota, S.; et al. Variance analysis of a clinical pathway of video-assisted single lobectomy for lung cancer. Surg. Today 2009, 39, 104–109. [CrossRef] [PubMed]

23. de Klundert, J.V.; Gorissen, P.; Zeemering, S. Measuring clinical pathway adherence. J. Biomed. Inform. 2010, 43, 861–872. [CrossRef] [PubMed] 24. Gath, I.; Geva, A.B. Unsupervised optimal fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 1989, 11, 773–780. [CrossRef]

24. 25. Lenard, M.J.; Alam, P. Application of fuzzy logic to fraud detection. In Encyclopedia of Information Science and Technology, 1st ed.; IGI Global: Hershey, PA, USA, 2005; pp. 135–139

ANNEXURES

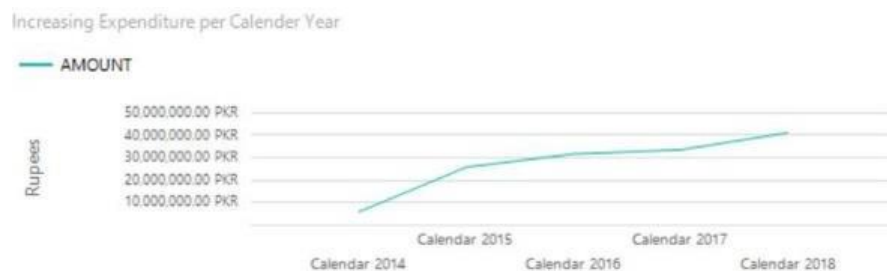
- Dataset Attributes:

- Unique number of servicesId: 1206
- Unique number of Doctors: 486
- Unique number of specialtyIds: 62
- Total number of transactions: 441,506

Attributes in Dataset	
Attributes Value Unique number of servicesId	1206
Unique number of Doctors	486
Unique number of specialtyIds	62
Total number of transactions	441,506

- Explanation of Figures:

- Fig-4.1 illustrates the volume of transactional data collected from insurance claim transactions over a span of five years.
- Fig- 4.2 depicts the functionality of the suggested fraud detection model, emphasizing the calculation of association scores for doctors and patients, followed by processing through a rule engine.
- Fig 4.3 presents the yearly increase in employee insurance coverage expenses from 2013 to 2019, indicating a trend of exponential growth possibly attributed to healthcare fraud.
- Fig 4.4 displays the percentages of improper payments in US federal health programs managed by the Health and Human Services (HHS) from 2012 to 2019, showing a continuous rise in improper payments under Medicaid and CHIP programs.
- The research primarily focuses on healthcare fraud detection methodologies using insurance claim transactional data from a local hospital over the period of 2013 to 2019. Publications addressing healthcare fraud in the United States are considered, and research with a holistic healthcare focus, where fraud identification is not the main goal, is excluded from this evaluation.



- The annexure provides additional information, data, and context related to the main content of the research report.
- It serves as a reference point for readers seeking detailed insights into the dataset,

figures, and scope of the research.