IJITCE

ISSN 2347- 3657

# International Journal of Information Technology & Computer Engineering


Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 15, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

# DETECTION OF FRAUD CLAIMS IN HEALTH INSURANCE INDUSTRY

*[1]Mr. V. Ranadheer Reddy,[2]R. Jhansi,[3]T. Bhavitha,[4]T. Deepthi*

*[1]Assistant Professor,[234]Students*

*Department Of CSE*

*Malla Reddy Engineering College for Women*

**Abstract:** For patients to pay for the expensive medical bills, they rely on health insurance offered by either private, public, or both systems. Some healthcare practitioners conduct insurance fraud as a result of their reliance on health insurance. Despite the tiny number of these service providers, it is said that fraud costs insurance companies billions of dollars annually. Our study involves formulating the fraud detection issue over minimum, definite claim data, which consists of procedure codes and medical diagnoses. Using a unique representation learning technique, we provide a solution to the fraudulent claim detection issue by converting procedure and diagnostic codes into Mixtures of Clinical Codes (MCC). We also look on ways to extend MCC using Robust Principal Component Analysis and Long Short Term Memory networks. Our test findings show encouraging results in the detection of false records.

## I. INTRODUCTION

### 1.1 Introduction

DATA analytics is becoming more and more important for almost every sector of economic growth. The vast quantity of data, including clinical data, prescription data, insurance claims, provider information, and patient information, "potentially" offers enormous opportunity for data analysts, as healthcare is one of the main financial sectors in the US economy. Every year, health insurance companies handle billions of claims, and the US spends more than $3 trillion on healthcare [1]. Using the many entities involved, Figure 1 provides a succinct flow of a typical healthcare reconciliation procedure. Before providing any services, the office of the service provider makes sure that the patient has sufficient coverage via his or her insurance plan or other finances. The service provider then uses the results of the first exams to determine pertinent diagnoses for the patient. After that, the patient is tested by the service provider utilizing one or more medical treatments, such as further diagnostic tests and surgical procedures. Along with additional data including personal, demographic, and visit history, these diagnoses and procedures are often associated with the patient's report. At this stage, the patient usually checks out and pays the copay specified by his or her insurance plan. After that, a medical coder receives the patient's report, abstracts the data, and prepares a "superbill" with all the provider's details on it. Seeing as how much money the healthcare sector brings in, it is not unusual to see falsified and fraudulent claims made to insurance companies. Healthcare fraud is defined as "an intentional deception or misrepresentation made by a person, or an entity, with the knowledge that the deception could result in some unauthorized benefit to him or some other entities" [3] by the National Health Care Anti-Fraud Association (NHCAA). Even if they only make up a tiny portion, such false claims come at a very expensive cost. The NHCAA estimates that the financial losses in the US due to fraud are in the tens of billions of dollars [3]. Studies reveal that a relatively little percentage of losses are recovered each year, despite the healthcare industry' stringent procedures addressing fraud and abuse control [4].

The following are the most frequent fraudulent acts carried out by dishonest healthcare practitioners.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

_ Making up diagnoses in order to support unnecessary medical treatments.
_ Upcoding, or billing for expensive procedures or services in lieu of the real procedures.
_ Making up claims for operations that were not completed.
_ Undertaking needless medical treatments in order to get insurance reimbursements.
_ "Unbundling," or billing for each phase of a treatment as if it were a stand-alone operation.
_ Making up claims that non-covered therapies are medically required in order to get insurance reimbursement, particularly for cosmetic operations.

Applying domain knowledge only to address all or a portion of the above-mentioned problems is neither practicable nor possible. Automated data analytics may be used to identify false claims early on and greatly assist subject matter specialists in better managing the fraudulent activity.

In this research, we address the issue of healthcare fraud detection from the perspective of health insurance companies. When we have limited data, such as diagnostic and procedure codes, we provide a response to the issue of how to categorize an operation as valid or fraudulent from a claim. Various techniques, including data mining [5], classification methods [6, 7], Bayesian analysis [8], statistical surveys [9], non-parametric approaches [10], and expert analysis, have been used to identify the fraud detection challenge in the medical sector. Current approaches construct models for claim status prediction based on information from a claim database such as the physician's profile, background history, claim amount, service quality, services done per provider, and associated indicators. These techniques work well, however they often need datasets that are not openly accessible. It is also very difficult to transfer the answers due to the diversity and general incompatibility of the variables included in such datasets. Because gaining third-party access to richer datasets is frequently forbidden by the Health Insurance Portability and Accountability Act (HIPAA) in the US, the General Data Protection Regulation (GDPR) in Europe, or similar laws in other regions, we have limited the data we have available for this study to diagnosis and procedure codes. In addition, compared to other industries, the healthcare sector is more reluctant to disclose data. Furthermore, it is not possible to transfer solutions from one software system to another since various software systems report different patient data. Consequently, we limit the scope of our issue formulation to diagnostic and procedure codes, which are universally manageable regardless of their national or international context. Our method of solving the problem is predicated on the assumption that the claim data is a combination of clinical codes for diagnoses and procedures in the International Classification of Diseases (ICD) coding system. Furthermore, the suggested method works flawlessly with other coding systems, such as Healthcare Common Procedure Coding System (HCPCS) and Current Procedural Terminology (CPT), or their mixtures.

Using probabilistic topic modeling, we describe an insurance claim as a Mixture of latent Clinical Concepts (MCC). To the best of our knowledge, this is the first study to use a latent space to describe insurance claims as combinations of clinical ideas. We take it for granted that each and every claim is an expression of hidden or apparent combinations of clinical concepts, including pain, mental, or infectious disorders. Additionally, every clinical concept consists of a combination of procedure and diagnostic codes. Our model's intuition is derived from the services offered by clinics, hospitals, and doctor's offices. Typically, a patient receives treatments in response to certain problems that may include one or more diagnoses. The service provider next treats the patient by carrying out the required procedures. As a result, a claim's diagnoses and procedures may be stated as a combination of clinical concepts, including

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

pain, mental, and/or infectious disorders, along with their respective treatments. Note that since these notions are often unclear, difficult, or need for specialized expertise, we do not specifically identify or interpret them.

We use Robust Principal Component Analysis and Long-Short Term Memory networks to enhance the MCC model. By expanding MCC, we want to separate the important ideas from assertions and categorize them as either false or not. By using a claim's concept weights as a sequence representation in an LSTM network, we expand on MCC. With the help of this network, we may express the claims as dependent idea sequences that the LSTM can classify. Similar to this, we break down claims into low-rank and sparse vector representations using Robust Principal Component Analysis (RPCA) in order to select relevant idea weights. The noise-free weights are best captured by the low-rank matrix.

We may sum up our distinctive contributions to this research as follows.

The challenge of detecting fraudulent claims is formulated using minimum and definite claim data, which consists of procedure and diagnostic codes.

Here, we provide a novel technique to representation learning: clinical ideas over procedure and diagnostic codes.

_ We use LSTM and RPCA for classification to broaden the combinations of clinical concepts.

We demonstrate enhanced performance when comparing our methods to a baseline technique and the Multivariate Outlier Detection (MOD) [11]. Two phases make up the Multivariate Outlier Detection approach, which is used to find unusual provider payments in Medicare claims data. To provide matching residuals, a multivariate regression model is first constructed using 13 manually selected characteristics. Next, an extended univariate probability model receives the residuals as inputs. In particular, they looked for potential outliers in the claim data using the probabilistic programming techniques in Stan [12]. With a modified issue formulation, the authors employ the identical CMS (Centers for Medicare and Medicaid Services) dataset that we use in our trials. While we use MOD on MCC characteristics, their research includes provider and beneficiary data pertaining to Medicare enrollees in the state of Florida. Conversely, a test claim is designated as the majority label in the training claim data by the baseline classifier.

Based on the inpatient dataset that we acquired from CMS, our experimental findings demonstrate that MCC + LSTM achieves accuracy, precision, and recall scores of 59%, 61%, and 50%, respectively. Furthermore, on the outpatient dataset, it shows accuracy, precision, and recall scores of 78%, 83%, and 72%, respectively. It is our belief that the framing of the issue, representation learning, and solution that have been given will spark fresh study on the use of minimum but conclusive evidence to identify false claims.

### 1.2 Problem Statement

Creating heuristics based on fraud indicators is the foundation of the conventional method for detecting fraud. One of two decisions on fraud would be made using these heuristics. Rules that specify whether a case has to be sent for inquiry would be framed in particular instances. In different circumstances, a checklist with ratings for the different fraud indications would be created. If the case has to be referred for investigation, it would be decided by adding together these scores and the claim's worth. Periodically, the thresholds and indicator selection criteria will be recalculated based on statistical testing.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

The problem with the aforementioned methods is that they mainly depend on physical intervention, which will have the following negative effects.

• Limited to using heuristic knowledge to work within a small range of known parameters, but acknowledging that certain additional traits may potentially influence choices

• Incapacity to comprehend contextual linkages that may not conform to the usual image between factors (such as region, client group, and insurance sales procedure). Industry experts' consultations reveal that there is no "typical model," making it difficult to identify the model that is appropriate for a given situation. • Model calibration is a manual process that must be carried out on a regular basis to account for evolving behavior and guarantee that the model takes feedback from investigations into account. It is difficult to do this calibration. There is a minimal rate of fraud in total claims.Generally speaking, 1% or fewer of the claims are categorized. Furthermore, aggressive efforts must be made to detect new fraud schemes.

From the standpoint of standard statistics, they are difficult. As a result, insurers are beginning to consider using machine learning capabilities. The idea is to provide the algorithm with a wide range of data without making any assumptions about the items' relevancy. The goal is for the machine to create a model based on detected frauds that can be evaluated using a range of algorithmic approaches on these known frauds.

### 1.3 Objective

Investigate different machine learning strategies to increase detection accuracy in unbalanced data. To get better prediction performance, the effects of feature engineering, feature selection, and parameter tinkering are investigated.

The data will be divided into three sections as part of the procedure: training, testing, and cross-validation. A subset of the data will be used to train the algorithm, and a testing set will be used to fine-tune its parameters. The cross-validation set performance will be analyzed for this. To make sure that the outcomes are consistent, the top-performing models will next be evaluated for different random splits of the data.

The activity was carried out on ApolloTM, Wipro's Anomaly Detection Platform, which uses a mix of predictive machine learning algorithms and pre-established criteria to find anomalies in data. It is based on open source and has a manageable and customizable library of pre-built algorithms that allow for quick implementation. Three layers make up this big data platform, as seen below.

## II. LITERATURE SURVEY

Among the most important problems in the vast healthcare system are fraud and misuse. Apart from fraudulent activities, inadvertent mistakes in paperwork result in substantial losses of money, time, and workforce. Solutions to the issues of fraud, misuse, and mistake detection in the pharmaceutical, medical, and allied fields are put out in a number of publications in the literature. Using the process-mining framework and the clinical pathways idea, Yang and Hwang created a fraud detection model that can identify frauds in the healthcare industry [13]. The approach makes use of a module that extracts structural patterns from input clinical cases, both good and negative. The program extracts the most common patterns from each clinical event. Next, a filtered dataset with labeled features is produced using

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

a feature-selection module. Lastly, the feature set is used to construct an inductive model that assesses new claims. Principal component analysis, association analysis, and clustering are all used in their process. The method was used using actual data obtained from Taiwan's National Health Insurance (NHI) program. The importance of the many elements that the writers used to produce patterns for both abusive and normal claims is not addressed. A predictive algorithm for identifying fraud and abuse was introduced by Bayerstadler et al. [14] utilizing manually labeled claims as training data. The approach uses a probability distribution for new claim invoices in order to forecast the fraud and abuse score. In particular, the authors suggested using latent variables to condense the representation patterns in medical claims using a Bayesian network. The likelihood scores for numerous fraud incidents are predicted in the prediction stage using a multinomial variable modeling. They also used Markov Chain Monte Carlo (MCMC) to estimate the model parameters [15]. A Medicare fraud detection system based on anomaly detection [17] was suggested by Zhang et al. [16]. The first component of the suggested approach is an algorithm based on spatial density, which is said to be more appropriate for medical insurance data than local outlier variables. Regression analysis is used in the second stage of the procedure to find the linear connections between various variables. The authors also noted that the method's applicability to newly arriving data is restricted. In order to detect fraud and abuse related legal situations in the healthcare industry, Kose et al. [18] employed interactive unsupervised machine learning, where expert knowledge is used as an input to the algorithm. Analytic Hierarchical Process (AHP) pairwise comparison approach was used by the authors to add weights between actors (i.e., patients) and characteristics. EM, or expectation maximization, is used to group agents that are similar. They created storyboards based on deviant behavior features and included domain experts in the research at various stages. The storyboard is used to identify behavior features, and the suggested framework is then reviewed. All associated individuals and goods, such pharmaceuticals, are included in the prescription process. In order to screen healthcare claims, Bauder and Khoshgoftaar [19] devised a universal outlier identification methodology using Bayesian inference. In their investigations, they used the Stan model, which is comparable to [20]. Be aware that they only take provider level fraud detection into account; clinical code-based relationships are not taken into account. Numerous techniques make use of either proprietary datasets or disparate datasets with incompatible feature lists. It is thus exceedingly challenging to directly compare these findings. Furthermore, healthcare providers and insurance companies are highly hesitant, if not unwilling at all, to exchange rich datasets due to the enforcement of severe fines for breaches of the privacy and security of healthcare information under HIPAA, GDPR, and similar laws. For these reasons, we frame the issue in terms of a basic, final claim set that consists of operation and diagnostic codes. In this context, we address the issue of classifying an operation as authentic or fraudulent by combining RNN and RPCA based encodings with combinations of clinical codes.

## III. SYSTEM ANALYSIS

### 3.1.1 Existing System

Using the process-mining framework and the clinical pathways idea, Yang and Hwang created a fraud detection model that can identify frauds in the healthcare industry [13]. The approach makes use of a module that extracts structural patterns from input clinical cases, both good and negative. The program extracts the most common patterns from each clinical event. Next, a filtered dataset with labeled features is produced using a feature-selection module. Lastly, the feature set is used to construct an inductive model that assesses new claims. Principal component analysis, association analysis, and clustering are all

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

used in their process. The method was used using actual data obtained from Taiwan's National Health Insurance (NHI) program. The importance of the many elements that the writers used to produce patterns for both abusive and normal claims is not addressed.

A predictive algorithm for identifying fraud and abuse was introduced by Bayerstadler et al. [14] utilizing manually labeled claims as training data. The approach uses a probability distribution for new claim invoices in order to forecast the fraud and abuse score. In particular, the authors suggested using latent variables to condense the representation patterns in medical claims using a Bayesian network. The likelihood scores for numerous fraud incidents are predicted in the prediction stage using a multinomial variable modeling. They also used Markov Chain Monte Carlo (MCMC) to estimate the model parameters [15].

A Medicare fraud detection system based on anomaly detection [17] was suggested by Zhang et al. [16]. The first component of the suggested approach is an algorithm based on spatial density, which is said to be more appropriate for medical insurance data than local outlier variables. Regression analysis is used in the second stage of the procedure to find the linear connections between various variables. The authors also noted that the method's applicability to newly arriving data is restricted.

In order to detect fraud and abuse related legal situations in the healthcare industry, Kose et al. [18] employed interactive unsupervised machine learning, where expert knowledge is used as an input to the algorithm. Analytic Hierarchical Process (AHP) pairwise comparison approach was used by the authors to add weights between actors (i.e., patients) and characteristics. EM, or expectation maximization, is used to group agents that are similar. They created storyboards based on deviant behavior features and included domain experts in the research at various stages. The storyboard is used to identify behavior features, and the suggested framework is then reviewed. All associated individuals and goods, such pharmaceuticals, are included in the prescription process.

In order to screen healthcare claims, Bauder and Khoshgoftaar [19] devised a universal outlier identification methodology using Bayesian inference. In their investigations, they used the Stan model, which is comparable to [20]. Be aware that they only take provider level fraud detection into account; clinical code-based relationships are not taken into account. Numerous techniques make use of either proprietary datasets or disparate datasets with incompatible feature lists. It is thus exceedingly challenging to directly compare these findings. Furthermore, healthcare providers and insurance companies are highly hesitant, if not unwilling at all, to exchange rich datasets due to the enforcement of severe fines for breaches of the privacy and security of healthcare information under HIPAA, GDPR, and similar laws. For these reasons, we frame the issue in terms of a basic, final claim set that consists of operation and diagnostic codes. In this context, we address the issue of classifying an operation as authentic or fraudulent by combining RNN and RPCA based encodings with combinations of clinical codes.

**Disadvantages**

1. Making up diagnoses in order to support treatments that are not required by medicine.
2. Making up claims for operations that were not completed.
3. Undertaking needless medical treatments in order to get insurance benefits.
4. "Unbundling," or billing for each phase of a process as if it were a distinct operation.
5. Falsifying non-covered treatments—especially cosmetic procedures—as medically essential in order to get insurance payments.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 15, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

**PROPOSED SYSTEM**

We use Robust Principal Component Analysis and Long-Short Term Memory networks to enhance the MCC model. By expanding MCC, we want to separate the important ideas from assertions and categorize them as either false or not. By using a claim's concept weights as a sequence representation in an LSTM network, we expand on MCC. With the help of this network, we may express the claims as dependent idea sequences that the LSTM can classify. Similar to this, we break down claims into low-rank and sparse vector representations using Robust Principal Component Analysis (RPCA) in order to select relevant idea weights. The noise-free weights are best captured by the low-rank matrix.

We may sum up our distinctive contributions to this research as follows.

The method uses basic, definite claim data, which consists of procedure and diagnostic codes, to construct the fraudulent claim detection issue.

The system presents a novel representation learning strategy by prioritizing clinical ideas above procedure and diagnostic data.

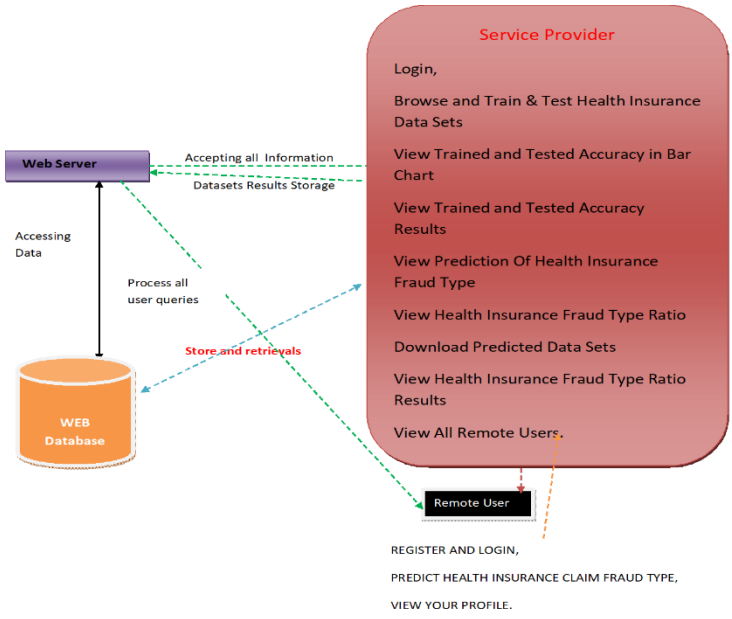
The system uses LSTM and RPCA for classification, extending combinations of clinical notions.

**Advantages**

- The suggested method classifies data using MCC by using Support Vector Machines (SVM).
- The Multivariate Outlier Detection technique is a useful tool for identifying unusual provider payments in Medicare claim data.

## IV. SYSTEM DESIGN

### 4.1 System Architecture




## V. SYSTEM IMPLEMENTATION

### Modules

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

**Modules Description**

**Service Provider**

The Service Provider must provide a valid user name and password to log in to this module. Following a successful login, one may do a number of tasks, including Examine, Train, and Test Data Sets for Health Insurance View the results of the trained and tested accuracy, view the bar chart showing the accuracy, view the prediction of the kind of health insurance fraud, view the ratio of the type of health insurance fraud, and download the predicted data sets. View All Remote Users and the Results of the Health Insurance Fraud Type Ratio

**View and Authorize Users**

The administrator may see a list of all enrolled users in this module. The administrator may see user information here, including name, email address, and address, and they can also approve people.

**Remote User**

There are n numbers of users present in this module. Prior to beginning any actions, the user must register. The user's information is saved in the database when they register. Upon successful registration, he must use his permitted user name and password to log in. After successfully logging in, the user may do several tasks such as registering and logging in, predicting health insurance fraud type, and seeing their profile.

## VI. RESULTS

**Dataset Description**

The author of this project divided it into two sections. In the first, he provided a brief overview of the literature on technologies that can be used to enhance vineyard growth. In the second, he described the "Health Clime -ML Database," which can be used to train different machine learning algorithms like SVM and logistic regression. The taught model may be used to forecast grape growth, harvest timing, and phenology (development cycle type) using fresh test photos after it has been trained using machine learning methods.

The creator of the database provided five distinct dataset kinds, which are explained below.

1) Dataset 1: Using this dataset, machine learning algorithms may be trained, and the resulting model can be used to forecast harvest time.

2) Dataset 2: This dataset may be used to train machine learning algorithms that forecast growth

3) Dataset 3: This is useful for forecasting the phenology stage.

4) Maturity may be predicted using datasets 4 and 5.

Note: We are unable to create that module since the dataset does not include any photos for illness prediction.

In this project, we use the first three datasets to forecast the harvest time, growth rate, and phenology type. We omit datasets 4 and 5 since they need too much processing time owing to their large picture sizes, and we have only used the SVM and logistic regression algorithms for the same reason.

The "Health Clime Database" folder, which has three separate datasets for gathering photos, growth rate, and phenology type, is where you can see all of the photographs. Screenshots of the dataset photos are below.

## VII. CONCLUSION

We present the issue of identifying fraudulent insurance claims as a feature generation and classification process in this project. As a result of legal restrictions and software system discrepancies, we construct

ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

the issue over a minimum, definite claim data set that consists of procedure and diagnostic codes. As a novel representation learning technique, we offer clinical ideas over procedure and diagnostic codes. Every claim is interpreted as a latent or evident blend of clinical concepts, which are themselves combinations of procedure and diagnostic codes. We expand the MCC model by filtering the important ideas from claims and categorizing them as fraudulent or non-fraudulent using the Long-Short Term Memory Network (MCC + LSTM) and Robust Principal Component Analysis (MCC + RPCA). Our findings show that there is room for improvement in the identification of fraudulent healthcare claims using limited data. In the negative claim generation process, both MCC and MCC + RPCA behave consistently for different concept sizes and replacement probabilities. Using the inpatient dataset, MCC + LSTM achieves accuracy, precision, and recall scores of 59%, 61%, and 50%, respectively. Additionally, on the outpatient dataset, it displays accuracy, precision, and recall scores of 78%, 83%, and 72%, respectively. The findings of MCC and MCC + RPCA exhibit similarities as they both make use of an SVM classifier. We think that the design of the suggested issue, representation learning, and solution will start fresh research on the use of minimum but conclusive data to identify fraudulent insurance claims.

**REFERENCES**

1. National Health Care Anti-Fraud Association, "The challenge of health care fraud," https://www.nhcaa.org/resources/health-care-antifraud-resources/the-challenge-of-health-care-fraud.aspx, 2020, accessed January, 2020.
2. Font Awesome, "Image generated by free icons,"https://fontawesome.com/license/free, 2020, online.
3. National Health Care Anti-Fraud Association, "Consumer info and action," https://www.nhcaa.org/resources/health-care-anti-fraudresources/ consumer-info-action.aspx, 2020, accessed January, 2020.
4. W. J. Rudman, J. S. Eberhardt, W. Pierce, and S. Hart-Hester, "Healthcare fraud and abuse," Perspectives in Health Information Management/ AHIMA, American Health Information Management Association, vol. 6, no. Fall, 2009.
5. M. Kirlidog and C. Asuk, "A fraud detection approach with data mining in health insurance," Procedia-Social and Behavioral Sciences, vol. 62, pp. 989–994, 2012.
6. V. Rawte and G. Anuradha, "Fraud detection in health insurance using data mining techniques," in 2015 International Conference on Communication, Information & Computing Technology (ICCICT). IEEE, 2015,pp. 1–5.
7. C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: classification of skewed data," Acm sigkdd explorations newsletter, vol. 6, no. 1, pp. 50–59, 2004.
8. T. Ekina, F. Leva, F. Ruggeri, and R. Soyer, "Application of Bayesian methods in detection of healthcare fraud," chemical engineering Transaction,vol. 33, 2013.
9. J. Li, K.-Y. Huang, J. Jin, and J. Shi, "A survey on statistical methods for health care fraud detection," Health care management science, vol. 11, no. 3, pp. 275–287, 2008.
10. R. J. Freese, A. P. Jost, B. K. Schulte, W. A. Klindworth, and S. T. Parente, "Healthcare claims fraud, waste and abuse detection system using non-parametric statistics and probability based scores," Jan. 19 2017, uS Patent App. 15/216,133.
11. R. A. Bauder and T. M. Khoshgoftaar, "Multivariate anomaly detection in medicare using model residuals and probabilistic programming," in The Thirtieth International Flairs Conference, 2017.
12. B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," Journal of statistical software,vol. 76, no. 1, 2017.
13. W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud

ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp791-800

and abuse," Expert Systems with Applications, vol. 31, no. 1, pp. 56–68, 2006.
14. A. Bayerstadler, L. van Dijk, and F.Winter, "Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance," Insurance: Mathematics and Economics, vol. 71, pp. 244–252, 2016.
15. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing markov chain monte carlo," Markov chain Monte Carlo in practice, vol. 1, p. 19, 1996.