



## International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

# Fraud Detection in Health Care Insurance Sector Using Datamining Upcoding Technique

P. Sailaja , Ilakkiya.V , Suji.D

Department of Information Technology, Velammal Institute of Technology, Chennai, Tamil Nadu, India

**ABSTRACT:** Statistics have shown that millions of dollars used on the healthcare expenditures annually are exhausted due to frauds. Within the healthcare sector, data mining technology plays a vital role in detection of frauds in insurance claims. Fraud detection is a scenario applicable to many industries such as banking and financial sectors, insurance, healthcare, government agencies and law enforcement and more. Upcoding fraud is one such fraud in which a service provider acquires additional financial gain by coding a service by upgrading it even though the lesser service has been performed. This paper gives an insight into the various data mining tools which are efficient in detecting upcoding frauds especially in the healthcare insurance sector in India.

**KEYWORDS:** Data Mining, Upcoding, Fraud Detection, Healthcare

### I. INTRODUCTION

Health care system plays an important role in the quality of life and in social welfare of the modern society. One of the important sectors in the healthcare systems is the healthcare insurance sector. Different forms of frauds create threats for mankind, be it social or financial sector. In banking, fraud may be using stolen credit cards, falsifying cheques, misleading accounts and more. In insurance, 20% to 25% of claims contain some form of fraud, leading to approximately 10% of insurance payout dollars. Fraud is an adaptive crime, so it requires special methods of intelligent data analysis to detect and prevent it. Fraud detection in health insurance uses data mining techniques. Fraud is widespread and very costly to the healthcare insurance system. Fraud involves intentional deception or misrepresentation intended to result in an unauthorized benefit. In order to detect and avoid the fraud, data mining techniques are applied.

Data analysis techniques to prevent fraud were first used by the telephone companies, the insurance companies and the banks. These methods exist in the areas of Knowledge Discovery in Databases (KDD), Data Mining, Machine Learning and Statistics. They offer applicable and successful solutions in different areas of fraud crimes.

### II. UPCODING

Upcoding is a fraudulent medical billing activity in which a healthcare provider bills an insurance provider, say a public or a private insurance company with a CPT (Current Procedural Terminology) code for a more expensive service than was performed. Upcoding is illegal and a fraudulent activity followed by medical providers by cheating the insurance providers and gain more income than they are eligible. Medicare is a government agency that provides medical services to seniors. Medical providers apply to be a part of the Medicare reimbursement system, and when those providers serve senior patients that have Medicare as their insurance company, Medicare will then pay for services rendered. When a provider performs a service and then sends a bill to Medicare for a higher paying service than what was performed, that is called upcoding. A provider billing Medicare understands the complex system of coding reimbursement. Most claims submitted by providers are computerized, and there are many companies that contract with Medicare to pay the claims using government funds. Certain ICD-10 codes are listed in the table below which are used widely by the insurer to settle claims.



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

Serial No	2017-ICD-10-Codes	Description
1	A00-B99	Certain infectious and parasitic diseases.
2	C00-D49	Neoplasms.
3	D50-D89	Diseases of the blood-forming organs.
4	E00-E89	Endocrine, nutritional.
5	F01-F99	Mental, Behavioral.
6	G00-G99	Nervous system
7	H00-H59	Eyes, Adnexa
8	H60-H95	Ear, Mastoid
9	I00-I99	Circulatory
10	J00-J99	Respiratory

### III. DATA MINING IN FRAUD DETECTION

Finding new insights from the existing one is known as mining of data. The existing data should satisfy some features such as relevance, availability, cleanliness and adequacy.

With the support of artificial intelligence and machine learning, fraud detection has become much simpler and easier and has been incorporated by business organizations to detect anomalies which would otherwise be a tedious and time consuming manual procedure.

#### Supervised Learning Methods

Supervised learning is a method in which an algorithm is applied between an input variable(x) and an output variable(y) to create a mapping function from the input to the output.

$$Y=f(x)$$

The goal is to approximately map the function efficiently that when a new input variable(x) is given it can easily predict the output variables(y) or that information.

#### Unsupervised Learning Methods

Unsupervised learning is a methodology where only the input variable(x) is known with no corresponding output variable. The goal of unsupervised learning is to mould an underlying structure in the data in order to learn mine truths about the data. This method typically analyses one claim's attributes in relation to other claims and discover how they are related to or differ from each other.

#### Hybrid or Semi-Supervised Learning Methods

This is a methodology in which there are huge amount of data(x) and only some of the data(y) are labeled. The actual problem resides in between supervised learning. A study followed a three step procedure for insurance fraud detection. It applied unsupervised clustering methods on insurance claims and modeled a variety of labeled clusters.

### IV. UPCODING RECOGNITION

Prior to the description of the earlier works involves the recognition of upcoding. After the person gets admitted to hospital, he is monitored and the services are provided. The physician records this in the medical chart. Now a medical coder assesses the data and translates them into the corresponding diagnosis using either ICD-9 or ICD-10 and procedure groups such as CPT. Diagnosis – Related Groups (DRGs) are formed by the combination of these diagnosis and groups.

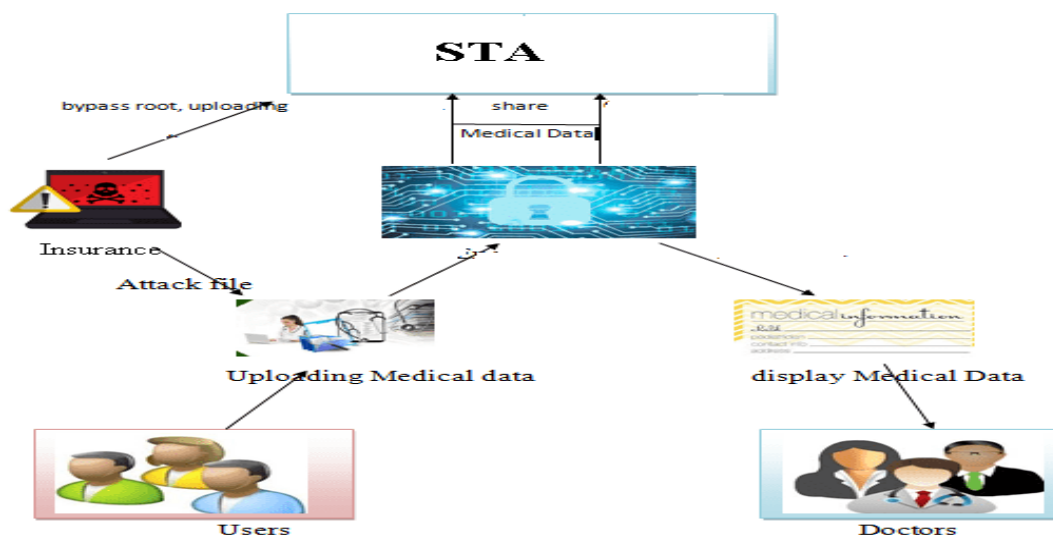
# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

## V. ARCHITECTURE DIAGRAM



### USER:

The Users major role is Registration and login. User need to create a separate account to them. This module authorize user into the system. This adds security to the use data. The login credentials are secured by encryption and they are decrypted back by the server to avoid eavesdropping. Once user logged in then the user can enter his/her query to the search query box. The entered query will stored into the user database.

### PROVIDER:

The provider module authorize as an admin into the system. The provider is the one who is having the all user details and hospital as well as insurance details. This module is used to verify the user details with health care issues. The provider can have all the user details with their health care issue. So that provider can search the results of the user query easily.

### STA:

STA is the third party generator between user and provider. STA is the one who can generate the token for all the users. The token generation can be done with the help of provider. (STA can verify user details as well as health care with the help of provider).

### HEALTHCARE DATA:

The health care fraud detection come mostly from insurance carriers including governmental health departments and private insurance companies. Major governmental health departments that have been reported in the literature include the Bureau of National Health Insurance (NHI) in Taiwan. The data from private insurance companies have also been used by several researchers.

## VI. CONCLUSION

Fraud including upcoding puts a heavy financial burden not only to the insurer but also to the customer as this would increase the premium rates payable to the insurer. So efficient mechanism has to be formulated and implemented to detect these financial frauds as well as to mitigate it. A healthcare fraud detection study is limited using supervised and unsupervised learning methodologies and is very few in the upcoding healthcare fraud detection. Bayesian models



ISSN(Online): 2320-9801  
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

are the supervised techniques used for upcoding fraud detection. A combination of subgroup creation via decision tree and Fisher's Extract Test are done using the unsupervised learning techniques.

## REFERENCES

- [1] Saba Kareem | Rohiza Binti Ahmad | Aliz Binit Sarlan "Framework for the identification of fraudulent health insurance claims using association rule mining." Conference of Big Data and Analytics (ICBDA) 2017.
- [2] Aryya Gangopadhyay | Song Chen "Health Care Fraud Detection with Community Detection Algorithms." International Conerence on Smart Computing (SMARTCOMP) 2016.
- [3] Sasongko, Nanang. "Fraud Detection In Non-Cash Transactions Using Information System Audit (A Case Study Of Government Bank in Bandung Indonesia)." ASEAN/Asian Asian Academic Society International Conference Proceeding Series. 2015.
- [4] Hsia, Renee Y., and Yaa Akosa Antwi. "Variation in charges for emergency department visits across California." Annals of emergency medicine 64.2 (2014) : 120-126
- [5] Jurges, Hendrik, and Juliane Koberlein-Neu. "First Do No Harm-Then Do Not Cheat: DRG Upcoding in German Neonatology." (2013).
- [6] Sofianita Mutalib, Rohayu Abd. Razak, Sharifalillah Nordin , Shuzlina Abdul Rahman , Azlinah Mohamed "Intelligent Classification in medical data." EMBS Conference n Biomedical Engineering and Sciences. 2012.
- [8] Phua, Cliton, et al. "A comprehensive survey of data mining- vased fraud detection research." arXic preprint arXiv:1009.6119 (2010).