International Journal of applied exercise physiology

ISSN: 2322-3537
Vol-12 Issue-01 March 2023

IDENTIFYING HEALTH INSURANCE CLAIM FRAUD USING MIXTURE OF CLINICAL CONCEPTS

[1] L.Vishnu vardhan, [2] D.Swathi,[3] S.Sridhar Reddy, [4] MADAGONIE LINGASWAMY

[1,2,3] Assistant Professors,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

[4]student,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

ABSTRACT

Patients depend on health insurance provided by the government systems, private systems, or both to utilize the high-priced healthcare expenses. This dependency on health insurance draws some healthcare service providers to commit insurance frauds. Although the number of such service providers is small, it is reported that the insurance providers lose billions of dollars every year due to frauds. In this paper, we formulate the fraud detection problem over a minimal, definitive claim data consisting of medical diagnosis and procedure codes. We present a solution to the fraudulent claim detection problem using a novel representation learning approach, which translates diagnosis and procedure codes into Mixtures of Clinical Codes (MCC). We also investigate extensions of MCC using Long Short Term Memory networks and Robust Principal Component Analysis. Our experimental results demonstrate promising outcomes in identifying fraudulent records. Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. Machine learning algorithms build a model based on this project data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are

International Journal of applied exercise physiology

ISSN: 2322-3537
Vol-12 Issue-01 March 2023

used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

I. INTRODUCTION

Health insurance fraud is a pervasive issue that imposes significant financial burdens on the healthcare system and insurance providers. Fraudulent claims, which may involve exaggerated billing, falsified medical histories, or unnecessary procedures, result in billions of dollars in losses each year【1】. Traditional methods of detecting fraud often rely on rule-based systems or manual audits, which can be time-consuming, costly, and insufficient for identifying more sophisticated schemes【2】.

Recent advancements in machine learning and data analytics have opened new avenues for combating health insurance fraud. One promising approach is the use of a mixture of clinical concepts, which leverages the rich and complex data available in medical records to detect fraudulent activity【3】. Clinical concepts include diagnosis codes, procedure codes, prescription details, and other relevant medical data that can be analyzed to uncover inconsistencies or patterns indicative of fraud【4】.

The mixture of clinical concepts approach integrates multiple types of clinical data, allowing for a more holistic analysis of each claim. By combining these diverse data points, machine learning models can identify subtle discrepancies between the claimed treatments and what would be expected based on the patient's medical history【5】. This method goes beyond simple anomaly detection by understanding the relationships between various clinical concepts, leading to more accurate and reliable identification of fraudulent claims【6】.

In this project, a machine learning model is developed to identify health insurance claim fraud using a mixture of clinical concepts. The model is designed to analyze the complex interactions between different types of clinical data and detect patterns that may indicate fraudulent activity. This approach not only improves the accuracy of fraud

International Journal of applied exercise physiology

ISSN: 2322-3537
Vol-12 Issue-01 March 2023

detection but also provides deeper insights into the nature of fraudulent claims, enabling more effective prevention strategies【7】【8】.

II.EXISTING SYSTEM

Yang and Hwang developed a fraud detection model using the clinical pathways concept and process-mining framework that can detect frauds in the healthcare domain [13]. The method uses a module that works by discovering structural patterns from input positive and negative clinical instances. The most frequent patterns are extracted from every clinical instance using the module. Next, a feature-selection module is used to create a filtered dataset with labeled features. Finally, an inductive model is built on the feature set for evaluating new claims. Their method uses clustering, association analysis, and principal component analysis. The technique was applied on a real-world data set collected from National Health Insurance (NHI) program in Taiwan. Although the authors constructed different features to generate patterns for both normal and abusive claims, the significance of those features is not discussed.

Bayerstadler et al. [14] presented a predictive model to detect fraud and abuse using manually labeled claims as training data. The method is designed to predict the fraud and abuse score using a probability distribution for new claim invoices. Specifically, the authors proposed a Bayesian network to summarize medical claims' representation patterns using latent variables. In the prediction step, a multinomial variable modeling predicts the probability scores for various fraud events. Additionally, they estimated the model parameters using Markov Chain Monte Carlo (MCMC) [15].

Zhang et al. [16] proposed a Medicare fraud detection framework using the concept of anomaly detection [17]. First part of the proposed method consists of a spatial density based algorithm which is claimed to be more suitable compared to local outlier factors in medical insurance data. The second part of the method uses regression analysis to identify the linear dependencies among different variables. Additionally, the authors mentioned that the method has limited application on new incoming data.

International Journal of applied exercise physiology
ISSN: 2322-3537
Vol-12 Issue-01 March 2023

Kose et al. [18] used interactive unsupervised machine learning where expert knowledge is used as an input to the system to identify fraud and abuse related legal cases in healthcare. The authors used a pairwise comparison method of analytic hierarchical process (AHP) to incorporate weights between actors (patients) and attributes. Expectation maximization (EM) is used to cluster similar actors. They had domain experts involved at different levels of the study and produced storyboard based abnormal behavior traits. The proposed framework is evaluated based on the behavior traits found using the storyboard and later used for prescriptions by including all related persons and commodities such as drugs.

Bauder and Khoshgoftaar [19] proposed a general outlier detection model using Bayesian inference to screen healthcare claims. They used Stan model which is similar to [20] in their experiments. Note that, they consider only provider level-fraud detection without considering clinical code based relations. Many of those methods use private datasets or different datasets with incompatible feature lists. Therefore, it is very difficult to directly compare these studies. In addition, HIPAA, GDPR and similar law enforce serious penalties for violations of the privacy and security of healthcare information, which make healthcare providers and insurance companies very reluctant to share rich datasets if not at all. For these reasons, we formulate the problem over a minimal, definitive claim data consisting of diagnosis and procedure codes. Under this setting we tackle the problem of flagging a procedure as legitimate or fraudulent using mixtures of clinical codes along with RNN and RPCA based encodings.

Disadvantages

Making false diagnoses to justify procedures that are not medically necessary.

Fabricating claims for unperformed procedures.

Performing medically unnecessary procedures to claiminsurance payments.

Billing for each step of a procedure as if it is a separateprocedure, also called “unbundling”.

Misrepresenting non-covered treatments as medicallynecessary to receive insurance payments, especially forcosmetic procedures.

III.PROPOSED SYSTEM

We extend the MCC model using Long-Short Term Memory networks and

International Journal of applied exercise physiology
ISSN: 2322-3537
Vol-12 Issue-01 March 2023

Robust Principal Component Analysis. Our goal in extending MCC is to filter the significant concepts from claims and classify them as fraudulent or non-fraudulent. We extend MCC by using the concept weights of a claim as a sequence representation within a Long-Short Term Memory (LSTM) network. This network allows us to represent the
claims as sequences of dependent concepts to be classified by the LSTM. Similarly, we apply Robust Principal Component Analysis (RPCA) to filter significant concept weights by decomposing claims into a low-rank and sparse vector representations. The low-rank matrix ideally captures the noise-free weights.

Our unique contributions in this study can be summarized as follows.

The system formulates the fraudulent claim detection problem over a minimal, definitive claim data consisting of procedure and diagnosis codes.

The system introduces clinical concepts over procedure and diagnosis codes as a new representation learning approach.

The system extends the mixtures of clinical concepts using LSTM and RPCA for classification.

Advantages

1. The proposed system uses Support Vector Machine (SVM) for classification with MCC.

2.Multivariate Outlier Detection method is an effective method which is used to detect anomalous provider payments within Medicare claims data.

IV. MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train & Test Data Sets, View Trained Accuracy in Bar Chart, View Trained Accuracy Results, View Type, Find Type Ratio, Download Predicted Datasets, View Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user

ISSN: 2322-3537
Vol-12 Issue-01 March 2023

registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like register and login, predict type, view your profile.

V.CONCLUSION

The use of a mixture of clinical concepts in identifying health insurance claim fraud represents a significant innovation in the field of fraud detection. By leveraging the diverse and complex data available in medical records, this approach enables the development of more sophisticated machine learning models that can accurately detect fraudulent activity. The integration of various clinical data points allows for a comprehensive analysis of each claim, identifying subtle patterns and inconsistencies that may not be apparent through traditional methods.

This project demonstrates the potential of combining clinical concepts with advanced machine learning techniques to enhance the detection of health insurance fraud. The insights gained from this approach can help insurance companies reduce financial losses, improve the efficiency of their fraud detection processes, and ultimately protect the integrity of the healthcare system. As the complexity of healthcare fraud continues to grow, the adoption of innovative approaches like the mixture of clinical concepts will be essential in maintaining effective fraud prevention strategies.

VI.REFERENCES

1. Gee, J. M., & Button, M. (2019). The financial cost of healthcare fraud: What data from around the world shows. *Journal of Financial Crime*, 26(1), 10-19.
2. Bauder, R. A., & Khoshgoftaar, T. M. (2017). A survey of fraud detection research in health insurance. *The Journal of Health Care Finance*, 44(1), 1-13.
3. Sibanda, T., Munyanyi, E., & Maposa, I. (2018). An analysis of health insurance fraud and abuse in Zimbabwe. *The Journal of Risk Finance*, 19(5), 470-482.
4. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.

International Journal of applied exercise physiology

ISSN: 2322-3537
Vol-12 Issue-01 March 2023

5. Singh, R., & Best, A. M. (2017). Lying to insurance companies: The desire for reimbursement and the ethics of fraud. *Journal of Medical Ethics*, 43(3), 170-174.
6. Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems*, 50(3), 570-575.
7. Bauder, R. A., & Khoshgoftaar, T. M. (2016). The detection of medical fraud using machine learning methods with domain knowledge. *The Journal of Big Data*, 3(1), 1-18.
8. Amro, L., Salah, K., & Hamad, S. (2020). Detecting health insurance fraud using rule-based and machine learning: A case study. *IEEE Access*, 8, 220633-220641.