The Use of a Combination of Clinical Concepts in the Determination of Health Insurance Claim Fraud

S JANAKI RAMUDU [1] -N PRABAKARAN [2], T MUNI SANKAR[3]

[1]Dep to computer science & Engineering , Manonmaniam Sundaranar University – Tirunelveli

2. Department of computer science & Engineering - A M Reddy Memorial college of engineering & Technology , Narsaraopeta

[3]Department of Computer science & Engineering , R K College of Engineering , Vijayawada

ABSTRACT

People rely on health insurance from either public or commercial systems, or both, to cover the expensive healthcare costs. As a result of their patients' reliance on health insurance, some medical professionals engage in insurance fraud. Insurance companies reportedly lose billions of dollars annually as a result of fraud, despite the fact that the number of such service providers is minimal. In this study, we provide a framework for detecting fraud using definite claim data that consists of medical procedure and diagnostic codes. By converting procedure and diagnostic codes into Mixtures of Clinical Codes (MCC), our innovative representation learning technique allows us to identify fraudulent claims. Long Short-Term Memory networks and Robust Principal Component Analysis are being explored as potential expansions of MCC. The findings of our experiments show promise in detecting fake records. The rapidly expanding discipline of data science relies heavily on machine learning. This project makes use of statistical approaches to train various types of algorithms to produce predictions or classifications and to discover important insights. Ideally, these insights will influence critical growth indicators and drive decision making inside applications and enterprises. This project data is used by machine learning algorithms to create a model, which may then make judgments or predictions without being specifically trained to do so. When traditional algorithm development becomes too daunting or impractical, many datasets turn to machine learning algorithms.

I. INTRODUCTION

The healthcare system and insurance companies have a heavy financial burden due to the widespread problem of health insurance fraud. Falsified medical histories, inflated invoices, or other forms of billing fraud needless processes cause annual losses amounting to billions of dollars. For increasingly complex fraud schemes, the time-consuming and expensive rule-based systems or human audits used in the past are frequently not enough.

Novel approaches to preventing health insurance fraud have emerged as a result of recent developments in data analytics and machine learning. The use of a combination of clinical ideas is an encouraging strategy that makes use of the extensive and intricate data found in medical records in order to identify instances of fraudulent conduct. You may evaluate pertinent medical data, such as diagnostic codes, procedure codes, prescription information, and clinical concepts, to find trends or discrepancies that might indicate fraud.

By combining different kinds of clinical evidence, the combination of clinical concepts method provides a more complete picture of each claim. Machine learning algorithms may find small differences between the patient's medical history and the therapies that were claimed by merging these different types of data. More accurate and dependable identification of fraudulent claims is achieved by this technology, which goes beyond basic anomaly detection, by comprehending the connections between different therapeutic concepts.

Using a variety of clinical principles, this study develops a machine learning model to detect fraudulent health insurance claims. Using the intricate interplay between several forms of clinical data, the model can spot trends that might point to fraud. The accuracy of fraud detection is enhanced and the nature of fraudulent claims is better understood using this technique, leading to more effective preventive tactics.

II.EXISTING SYSTEM

Using the idea of clinical pathways and a process-mining framework, Yang and Hwang created a fraud detection model that can identify healthcare-related frauds [13]. Using both good and negative clinical examples as input, the method's module finds structural patterns. The module is designed to extract the most common patterns from each clinical occurrence. After that, a filtered dataset with labelled features is generated using a feature-selection module. Lastly, the feature set is used to build an inductive model that can evaluate new claims. Their approach makes use of PCA, clustering, and association analysis. To test the method, researchers in Taiwan used data from the country's National Health Insurance (NHI) program. The authors did not address the relevance of the attributes they built to produce patterns for abusive and regular claims. A predictive model for the detection of fraud and abuse was introduced by Bayerstadler et al. [14]. The model was trained using claims that had been manually labelled. With the use of a probability distribution, the approach may anticipate the fraud and abuse score for newly submitted claim invoices. In order to summarize the representation patterns of medical claims using latent variables, the authors specifically suggested a Bayesian network. Step one involves making predictions about the likelihood of different types of fraud using multinomial variable modeling. Furthermore, they used Markov Chain Monte Carlo (MCMC) [15] to estimate the model values.

A Medicare fraud detection system based on the idea of anomaly detection was suggested by Zhang et al. [16]. Part one of the suggested approach is an algorithm based on geographical density that is said to be more suited to medical insurance data than local outlier characteristics. The second step of the process involves finding the linear relationships between variables by using regression analysis. The authors also noted that the strategy isn't very useful for processing fresh data.

To find healthcare fraud and abuse situations, Kose et al. [18] used interactive unsupervised machine learning, which takes expert knowledge as input. In order to account for the relative importance of patients and their characteristics, the authors used a pairwise comparison technique from the analytic hierarchical process (AHP). To group agents that are similar, expectation maximization (EM) is used. Storyboards based on deviant behavior features were created with the help of domain specialists who were participating in the research at various stages. 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 separate procedure, also called "unbundling".

Misrepresenting non-covered treatments as medically necessary to receive insurance payments, especially for cosmetic procedures.

III.PROPOSED SYSTEM

Long-Short Term Memory networks and Robust Principal Component Analysis are used to expand the MCC model. Our objective in expanding MCC is to identify and categorize claims as either fraudulent or non-fraudulent based on the important ideas they include. We improve upon MCC by training an LSTM network to represent claims' idea weights as a sequence. We may use this network to express the claims as LSTM-classifiable sequences of dependent notions. By breaking claims down into sparse vector representations with low ranks, we may use Robust Principal Component Analysis (RPCA) to filter out concepts with substantial weights. Optimal weights free of noise are captured by the low-rank matrix.

We may highlight our distinctive contributions to this work as follows.

Using limited, definite claim data including procedure and diagnostic codes, the system formulates the fraudulent claim detection issue.

As a novel representation learning strategy, the system incorporates clinical ideas into procedure and diagnostic codes.

This system uses LSTM and RPCA to classify combinations of clinical ideas.

The benefits

1. Support Vector Machine (SVM) is used by the suggested system for classification purposes using MCC.

An excellent tool for detecting abnormal provider payments within Medicare claims data is the Multivariate Outlier Detection method.

IV. ELEMENTS

Supplier of Services

A valid username and password are required for the Service Provider to access this module. Login, Train and Test Data Sets, and Other Operations Will Be Available to Him Once He Has Successfully Logged In. View All Remote Users, Download Predicted Datasets, View Type, Find Type Ratio, View Trained Accuracy Results in a Bar Chart, and View Type in General.

Keep an eye on and approve users

This section allows the administrator to get a complete rundown of all registered users. Here, the administrator may see the user's information (name, email, and address) and grant them access.

Work from afar

All all, there are n users in this module. Registration is required prior to performing any operations. The user's information will be entered into the database after they register. After he has successfully registered, he will need to log in using the permitted credentials. After logging in, users will be able to do things like examine their profile, anticipate their typing type, and register and login.

V.CONCLUSION

A major step forward in the fight against fraud in health insurance claims is the use of a variety of therapeutic ideas. This method allows for the creation of more advanced machine learning models that can correctly identify fraudulent behavior by making use of the varied and intricate data found in medical records. A thorough evaluation of every claim is possible because to the integration of different clinical data points, which may reveal irregularities and subtle patterns that would otherwise go unnoticed using more conventional approaches.

Health insurance fraud may be better detected with the help of this experiment, which shows how clinical principles can be combined with powerful machine learning algorithms. Insurance firms may use the insights gathered from this technique to save costs, make their fraud detection procedures more efficient, and safeguard the healthcare system as a whole. Innovative techniques, such as the combination of clinical principles, will be crucial in sustaining successful fraud prevention tactics as the complexity of healthcare fraud continues to expand.

VI.REFERENCES

1. The authors are Gee and Button (2019). Healthcare fraud's monetary impact: what global data reveals. Publication: Journal of Financial Crime, Volume 26, Issue 1, pages 10–19.

Second, in 2017, Bauder and Khoshgoftaar published. An analysis of health insurance fraud detection studies. Journal of Health Care Finance, 44(1), 1-13.

It was published in 2018 by Sibanda, Munyanyi, and Maposa. An examination of the misuse and fraud around health insurance in Zimbabwe. Volume 19, Issue 5, Pages 470–482, Journal of Risk Finance.

In 2014, Raghupathi and Raghupathi published a paper. The potential and promise of big data analytics in healthcare. Journal of Health Information Systems, 2(1), 3.

5. Best, A. M., and Singh, R. (2017). Lying to insurance companies: The desire for repayment and the ethics of deception. The address is 43(3), 170-174, and the journal is Journal of Medical Ethics.

6. In 2011, Zhou and Kapoor published. Spotting fraudulent financial statements in an evolutionary context. Chapters 570–575 of Decision Support Systems, Volume 50, Issue 3.

7. Bauder & Khoshgoftaar (2016). 1. Machine learning techniques augmented with domain knowledge for the purpose of detecting medical fraud. Publication: The Journal of Big Data, Volume 3, Issue 1, Pages #1–18.

In 2020, Amro, Salah, and Hamad published a study. A case study on the use of rule-based and machine learning algorithms for the detection of health insurance fraud. 8(10), 220633-220641, IEEE Access.