International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 2 Feb. 2024, pp: 01-04 www.ijaem.net ISSN: 2395-5252

A Block Chain and Machine Learning Approach to Medical Insurance Fraud Detection

Suraj B S
Department of Computer Science and Engineering
PESCE,Mandya

PESCE,Mandya
Chetan kumar V
Department of Computer Science and Engineering
PESCE,Mandya

Sindhu B S
Department of Computer Engineering and Engineering
PESCE,Mandya

C Chethana
Department of Computer Science and Engineering

Date of Submission: 01-02-2024 Date of Acceptance: 10-02-2024

Abstract-The responsibility of the Health Insurance Department to regulate medical spending has grown in importance due to the notable increase in medical costs. Because traditional medical insurance settlements are paid on a per-service basis, many needless expenses are incurred. The single-disease payment system has been used a lot these days to deal with this problem. Nevertheless, single-disease payments carry the risk of fraud. The authors of this work have presented a methodology for identifying health insurance fraud that is ingrained in block chains and machine learning techniques such as logistic regression and support vector machine (SVM) that can automatically identify anxious medical records in order to ensure the sustainable execution of single-disease payment and lessen the workload of medical insurance workers. Additionally, the authors have suggested a method for managing and storing medical records based on collaborative blockchain to ensure data integrity, auditability, traceability, and security. Experiments on two real datasets from two 3A hospitals show that the proposed method may be able to detect fraud successfully and significantly improve the efficiency of medical insurance evaluations..

Keyword: Fraud Detection, Block Chain, Medical Insurance, Support Vector Machine (SVM), Logistic Regression,

I. INTRODUCTION

The rapid growth of medical information has led to an enormous quantity of data being gathered in hospital information systems, which is driving the medical sector's big data generation. Hospital big data has shown tremendous benefits to the medical area, drawing interest from both academics and industry. One of the main areas of big data study in the medical sciences is medical cost control. In the healthcare industry, costs are covered by the traditional insurance system according to hospital-provided services, which leads to excessive medical treatment and higher treatment costs. In order to overcome the issues of component-based pricing schemes, a single sickness Payment based on deployment and detection-oriented groups that have undergone extensive research. Every illness has a set price under the single sickness payment approach usual payment terms. The social healthcare insurance company reimburses hospital expenses according to the recommended criteria for the illness [1].
To have a better understanding of the differences between a payment system based on a single disease and one based on products, take a look at the example below.

If a patient is diagnosed with cancer, the standard health insurance settlement process will result in a recompensing fee of 160,000 rupees if the health insurance rates are 200,000 rupees and the recompensing ratio is 80%. If the cancer diagnosis-related group weighting rate in the Diagnosis Related Groups payment model is 7.4 and the unit price is 15,000 rupees, the health insurance must pay 87,800 rupees [2].

Medical assets can be used consistently; for example, the amount of medical facilities used correlates with the number of patients admitted, the severity of the sickness, and the quality of the services provided. In conclusion, by limiting overuse of medical care, the single-illness payment model establishes a clear health

insurance payment standard for each sickness. This lowers hospital costs. This strategy guarantees excellent medical care while keeping things user-friendly.

The chief complaint is the term used by the doctor to characterize a patient's ailments or symptoms, together with the length of the issues, in the patient's clinical report. The chief complaint serves as the report's main focus. In accordance with the writing standards for Chinese electronic medical records, this should clearly state the characteristics of the first disease identification and be concise, well-written, and accurate, including no more than 30 characters [3].

II. RELATED WORKS

The viability of the detected sickness needs to be verified based on the principal symptoms of the patients in order to reveal erroneous clinical reports. The objective of detecting medical records that are dangerous is converted into a text classification problem, wherein the authors predict the probability of each ICD 10 category by analyzing a significant complaint. The estimated likelihood of each ICD 10 code will then be listed in decreasing order. If the medical record's assigned ICD-10 code is in the top-k set of predicted results, it will be considered reasonable. If not, it'll be considered a fraud and need human auditing. Depending on the situation, the k value may be calculated; a greater

III. PROPOSED METHOD

The technique suggested in this research is pipeline-based. Prior to classifying these datasets using machine learning techniques, the authors had to first collect data from various sources, then use blockchain technology to keep medical records and management**.**

In antifraud investigations, primary complaints and ICD are essential pieces of evidence. For robust tamper resistance, anti repudiation, dependability, and incorruptibility protection, the authors offer a block chain based architecture for medical data administration and storage [6][7].

Three areas comprise the main contributions to the suggested framework:
The authors have enhanced the efficiency of block chain consensus and the agreement process by splitting the vertices of the block chain into application and agreement vertices.The authors have developed a system for storing and accessing medical data based on smart contracts that guarantees the validity, tractability, and security of the data.

Block chain technology controls the generation, influx, and use of trustworthy data by principles of transparency and trust [8]. Block chains are often maintained via peer-to-peer networks, and the data they store is remarkable, tractable, and immutable. In terms of technological architecture, block chain is a complete solution that includes game theory, smart contracts, peer-to-peer networks, distributed storage, block-chain data structures, consensus algorithms, and cryptography algorithms [9].
There are three primary types of block chain systems: public, private, and consortium. Anybody can engage in a public chain, even anonymously. An internal company-only block chain is called a private block chain.

The block chain's users include physicians, hospitals, insurance companies, and regulatory authorities. Electronic medical records are created by doctors [10]. The physician must digitally sign and start a transaction when uploading the medical record to the block chain. Following a review, the doctor's office will digitally sign the medical record in order to obtain block chain certification. When evaluating the accuracy of sickness diagnoses, the Medical Insurance Center examines block chain data and examines dubious medical documents. Regulatory agencies can connect to the block chain as a vertex or communicate with it via terminals to synchronize all information for inspection and tracking the universal trafficking process.

During physical audits, the health insurance center's staff scans the block chain for pertinent medical information and signatures, then links them to affiliated physicians and hospitals. The audit results will be signed by physicians, hospitals, and health insurance providers and stored on the block chain. Physicians and hospitals that engage in medical insurance fraud will face consequences, and their fraudulent activities will be listed on their credit records. The medical insurance center may impose further limitations on physicians and hospitals with poor credit. The writers must divide the datasets into two categories: one for fraud and another for non-fraud. SVM and logistic regression are utilized to carry out these binary classifications [12].

The authors of this study used a support vector that lacked a kernel function. As far as the authors are aware, SVM divides into two segments via the hyperplane. Both logistic regression and SVM models are fairly basic in their application and yield binary classifier output.

IV. RESULTS AND DISCUSSION

The authors used two hospital data sets, each including 19910 and 53430 patient records, from Medicals. Based on the study, only seven illnesses were found in each hospital data set. A new dataset was built utilizing block chain technology to determine whether or not patients had insurance or a card. The Python 3 programming language is used to create the implementations, with system specifications 8 GB RAM and i5 processor. The primary page of hospital databases was used to extract the ICD codes, and admission records were used to gather the majority of patient complaints. Expert coders have updated the ICD codes on the hospital report's main page to reflect all of the admitted patients' medical information, making them more accurate than the diagnostic codes that the doctors recorded when the patient was hospitalized.

V. CONCLUSION

This article primarily uses the block chain technique for health insurance fraud detection. Using label and character representations, it determines the validity of the ICD code submitted in the health report and predicts the likelihood of an illness based on the primary complaint of the unwell individual. The block chain technology used by the authors to store and manage medical data guarantees information security, dependability, immutability, traceability, and authentication. The recommended approach may lighten the workload for brokers of health insurance while also increasing effectiveness. Trials on real datasets from two reputable hospitals suggest that this strategy has good interpretability and may carry out health insurance antifraud effectively. The authors have utilized two machine learning models in their suggested work to goal of classification. SVC provides greater accuracy, or more than 95.8%, compared to both the model SVC and the logistic regression classifier.

REFERENCES

[1]. Mathauer, Inke, and Friedrich Wittenbecher. "Hospital payment systems based on diagnosis-related groups: experiences in low-and middle-income countries." Bulletin of the World Health Organization 91 (2013): 746-756A.
[2]. Gao, Feng, et al. "Discussion on the implementation of single disease payment." Chinese Medical Record English Edition 1.1 (2013): 8-10.
[3]. Viloria, Amelec, et al. "Diabetes diagnostic prediction using vector support machines." Procedia Computer Science 170 (2020): 376-381.
[4]. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
[5]. Gao, Weichao, William G. Hatcher, and Wei Yu. "A survey of blockchain: Techniques, applications, and challenges." 2018 27th international conference on computer communication and networks (ICCCN). IEEE, 2018.
[6]. Mohapatra, Srikanta Kumar, et al. "Text classification using NLP based machine learning approach." AIP Conference Proceedings. Vol. 2463. No. 1. AIP Publishing LLC, 2022.
[7]. kumar Mohapatra, Srikanta, et al. "Internet of Things: Security, Challenges, Open Problems & Solutions." 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2021.
[8]. Sahu, Premananda, et al. "Detection and Classification of Encephalon Tumor Using Extreme Learning Machine Learning Algorithm Based on Deep Learning Method." Biologically Inspired Techniques in Many Criteria Decision Making. Springer, Singapore, 2022. 285-295.
[9]. Bamunif, Abdullah Omar Ali. "Sports Information & Discussion Forum Using Artificial Intelligence Techniques: A New Approach." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.11 (2021): 2847-2854.
[10]. Verma, Akhil. "Encryption and decryption of images based on steganography and cryptography algorithms: a new model." Turkish Journal of Computer and

Mathematics Education (TURCOMAT) 12.11 (2021): 2839-2846.
[11]. Amponsah, Anokye Acheampong, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. "A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology." Decision Analytics Journal 4 (2022): 100122.
[12]. Ngai, Eric WT, et al. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature." Decision support systems 50.3 (2011): 559-569.
[13]. Lynam, Anita L., et al. "Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults." Diagnostic and prognostic research 4.1 (2020): 1-10.
[14]. Zhang, Guoming, et al. "Identifying fraud in medical insurance based on blockchain and deep learning." Future Generation Computer Systems 130 (2022): 140-154.
[15]. Kapadiya, Khyati, et al. "Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects." IEEE Access 10 (2022): 79606-79627.