Automatic Recommender System for Smart-Contracts-based Healthcare Insurance Fraud Detection Development Platform: Design, Implementation, and Performance Evaluation

Rima Kaafarani, Leila Ismail, Oussama Zahwe

Submitted to: Journal of Medical Internet Research
on: July 11, 2023

Table of Contents

Automatic Recommender System for Smart-Contracts-based Healthcare Insurance Fraud Detection Development Platform: Design, Implementation, and Performance Evaluation

Rima Kaafarani[1]; Leila Ismail[2] DPhil; Oussama Zahwe[1]

[1]ICCS-Lab, Computer Science Department, American University of Culture and Education Beirut LB
[2]Intelligent Distributed Computing and Systems (INDUCE) Laboratory, Department of Computer Science and Software Engineering, College of Information Technology, National Water and Energy Center, United Arab Emirates University, Abu Dhabi, United Arab Emirat Abu Dhabi AE

Corresponding Author:
Leila Ismail DPhil
Intelligent Distributed Computing and Systems (INDUCE) Laboratory, Department of Computer Science and Software Engineering, College of Information Technology, National Water and Energy Center, United Arab Emirates University, Abu Dhabi, United Arab Emirat
15551 Al Maqam Campus
Al Ain
Abu Dhabi
AE

Abstract

Background: Healthcare insurance fraud is on the rise in many ways, such as falsifying information and hiding third-party liability. This can result in significant losses for the medical health insurance industry. Consequently, fraud detection is crucial. Currently, companies employ auditors who manually evaluate records and pinpoint fraud. However, an automated and effective method is needed to detect fraud with the continually increasing number of patients seeking health insurance. Blockchain is an emerging technology among businesses and is constantly evolving to meet their needs. With its characteristics of immutability, transparency, traceability, and smart contracts, it demonstrated its potential in the healthcare domain. In particular, smart contracts are essential to reduce the costs associated with traditional methods, which are mostly manual, while preserving privacy and building trust among healthcare stakeholders, including the patient and the health insurance networks. However, with so many blockchain options available, selecting the right one for healthcare insurance can be difficult.

Objective: This paper aims to develop and implement smart contracts for detecting healthcare insurance fraud efficiently. Therefore, we provide a taxonomy of fraud scenarios and implement their detection using a blockchain platform that is suitable for healthcare insurance fraud detection. To automatically and efficiently select the best platform, we propose and implement a decision-map-based recommender system. For the aim of developing the recommender system, we propose a taxonomy of 102 blockchain platforms.

Methods: We developed and implemented smart contracts for 12 fraud scenarios that we identified in the literature. We used the two top blockchain platforms selected by our proposed decision-making map-based recommender system, which is tailored for healthcare insurance fraud. In addition, we present a taxonomy of 102 blockchain platforms classified according to the application domains for which they can be used.

Results: The developed decision-map-based recommender system demonstrates that Hyperledger Fabric is the best blockchain platform for identifying healthcare insurance fraud. We demonstrate the effectiveness of our recommender system by comparing the performance of the top two platforms selected by our system. The blockchain platforms taxonomy that we created for this revealed that 59 blockchain platforms are suitable for all application domains, 25 for financial services, and 18 for various application domains. We designed and implemented fraud detection based on smart contracts.

Conclusions: Our decision-map recommender system, which is based on our proposed taxonomy of 102 platforms, automatically selected the top two platforms, which are Hyperledger Fabric and Neo, for the implementation of healthcare insurance fraud detection. Our performance evaluation for the two platforms indicates that Fabric surpassed Neo in all performance metrics, as depicted by our recommender system. We provided an implementation of fraud detection based on smart contracts.

(JMIR Preprints 11/07/2023:50730)
DOI: https://doi.org/10.2196/preprints.50730

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Automatic Recommender System for Smart-Contracts-based Healthcare Insurance Fraud Detection Development Platform: Design, Implementation, and Performance Evaluation

Rima Kaafarani[1], MS; Leila Ismail[23], PhD; Oussama Zahwe[1], PhD
[1] ICCS-Lab, Computer Science Department, American University of Culture and Education, 1507 Beirut, Lebanon
[2] Intelligent Distributed Computing and Systems (INDUCE) Laboratory, Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Abu Dhabi, United Arab Emirates
[3] National Water and Energy Center, United Arab Emirates University, Abu Dhabi, United Arab Emirates

Corresponding Author:
Leila Ismail, PhD
Intelligent Distributed Computing and Systems (INDUCE) Laboratory, Department of Computer Science and Software Engineering, College of Information Technology
United Arab Emirates University
Abu Dhabi, United Arab Emirates
Phone: 971 37673333 ext. 5530
Email: leila@uaeu.ac.ae; leilaism@gmail.com

Highlights

- We propose an automatic decision-map-based recommender system, which maps available blockchain development platforms with their corresponding features and detects the most suitable platform for implementing blockchain-based solutions.
- We propose a taxonomy of 102 blockchain platforms based on their applications domains, which revealed that 59 blockchain platforms are suitable for all application domains, 25 for financial services, and 18 for various application domains.
- We implement an automatic decision-map-based recommender system for healthcare insurance fraud detection.
- We provide a taxonomy of fraud scenarios and implement their detection on the top two selected blockchain platforms, by our recommender system, that are suitable for healthcare insurance fraud detection.
- We design and implement smart contracts for detecting healthcare insurance fraud.
- We evaluate our recommender system by comparing the performance of the top two platforms selected by our recommender system, for healthcare insurance fraud detection.

Abstract

Background: Healthcare insurance fraud is on the rise in many ways, such as falsifying information and hiding third-party liability. This can result in significant losses for the medical health insurance industry. Consequently, fraud detection is crucial. Currently, companies employ auditors who manually evaluate records and pinpoint fraud. However, an automated and effective method is needed to detect fraud with the continually increasing number of patients seeking health insurance.

Blockchain is an emerging technology among businesses and is constantly evolving to meet their needs. With its characteristics of immutability, transparency, traceability, and smart contracts, it demonstrates its potential in the healthcare domain. In particular, self-executable smart contracts are essential to reduce the costs associated with traditional paradigms, which are mostly manual, while preserving privacy and building trust among healthcare stakeholders, including the patient and the health insurance networks. However, with the proliferation of blockchain development platform options available, selecting the right one for healthcare insurance can be difficult. This paper addresses this void and develops an automated decision-map recommender system to select the most effective blockchain platform for insurance fraud detection.

Objective: This paper aims to develop and implement smart contracts for detecting healthcare insurance fraud efficiently. Therefore, we provide a taxonomy of fraud scenarios and implement their detection using a blockchain platform that is suitable for healthcare insurance fraud detection. To automatically and efficiently select the best platform, we propose and implement a decision-map-based recommender system. For the aim of developing the recommender system, we propose a taxonomy of 102 blockchain platforms.

Methods: We developed and implemented smart contracts for 12 fraud scenarios that we identified in the literature. We used the two top blockchain platforms selected by our proposed decision-making map-based recommender system, which is tailored for healthcare insurance fraud. In addition, we present a taxonomy of 102 blockchain platforms classified according to the application domains for which they can be used.

Results: The developed decision-map-based recommender system demonstrates that Hyperledger Fabric is the best blockchain platform for identifying healthcare insurance fraud. We demonstrate the effectiveness of our recommender system by comparing the performance of the top two platforms selected by our system. The blockchain platforms taxonomy that we created for this revealed that 59 blockchain platforms are suitable for all application domains, 25 for financial services, and 18 for various application domains. We designed and implemented fraud detection based on smart contracts.

Conclusions: Our decision-map recommender system, which is based on our proposed taxonomy of 102 platforms, automatically selected the top two platforms, which are Hyperledger Fabric and Neo, for the implementation of healthcare insurance fraud detection. Our performance evaluation for the two platforms indicates that Fabric surpassed Neo in all performance metrics, as depicted by our recommender system. We provided an implementation of fraud detection based on smart contracts.

KEYWORDS

Blockchain development platform; Health insurance, fraud detection, smart contract; healthcare insurance, Taxonomy.

Introduction

Healthcare insurance fraud presents a significant challenge for both the medical industry and government bodies. It represents a serious concern for the insurance industry due to fraud's financial impact on policyholders and insurance companies. According to the National Health Care Anti-Fraud Association, healthcare fraud leads to the loss of tens of billions of dollars annually [CITATION 235 \l 1033]. In 2020, according to the U.S. Department of Justice, a noteworthy accomplishment in combating healthcare fraud recovering $2.7 billion through settlements and judgment was announced, however, representing a significant 50% increase compared to the previous year[CITATION USD23 \l 1033]. Furthermore, the global healthcare fraud analytics market demonstrates substantial growth, rising from $2.43 billion in 2022 to $3.09 billion in 2023, reflecting a Compound Annual Growth Rate of 27.0%[CITATION Hea23 \l 1033]. On the other hand, health insurance is crucial to ensure people’s lives due to the high cost of medical treatments. The advantage of health insurance is being threatened by theft and fraudulent claims. With the

increasing number of patients' demands for health insurance, manual auditing for validating and pinpointing fraud is no longer efficient. Therefore, it is essential to create an automatic and efficient system that automatically detects fraud.

Therefore, machine learning solutions for detecting frauds that rely on datasets to train models for fraud detection were introduced[CITATION Ala22 \l 1033][CITATION Roh21 \l 1033]. However, they raise ethical concerns as the trained models could be biased and leave behind the minority [CITATION Ire21 \l 1033], and privacy and security issues [CITATION KHY22 \l 1033] due to the potential compromise of sensitive Personally Identifiable Information (PII) of patients. These considerations would have severe consequences, including reputational damage to insurance firms. Machine learning models should rely on high-quality data[CITATION Din18 \l 1033]. Therefore, they are still not trustworthy so far. Recently, Blockchain has emerged as a decentralized technology to implement secure transactions in a peer-to-peer network. It consists of a series of interconnected blocks of transactions. Each block contains data and is secured through cryptographic measures, such as hash functions and public/private key encryption[CITATION Lei201 \l 1033]. Transactions occur between nodes in a peer-to-peer network, without the need for a central authority. All transactions are recorded in an immutable ledger, and peers can only add to the ledger, not alter, or delete any previously recorded information[CITATION Ism202 \l 1033]. When a new node joins the network, it downloads a copy of the ledger. Before adding a block to the blockchain, a consensus is reached among peers. In addition, blockchain can execute smart contracts[CITATION Ism191 \l 1033].

Blockchain demonstrated its potential in various domains, including the healthcare system [CITATION Lei20 \l 1033][CITATION Ism203 \l 1033]. In particular, smart contacts in a blockchain were introduced as self-executing agents based on the transactions being executed[CITATION Lei211 \l 1033]. However, there are proliferation of blockchain development platforms in the literature with various characteristics imposing challenges for software developers to determine suitable platforms which include the functionalities needed to implement insurance fraud detection solutions based on smart contracts. In this paper, we propose an automated decision-map recommender system specifically designed to select the most suitable blockchain platform among the proposed platforms in the literature and design. We exemplify the use of our proposed recommender system by implementing smarts-contract-based solutions for insurance fraud detection on the selected platform. The main contributions of this research work are as follows:

- We propose and develop an innovative and adaptive automated recommender system based on our proposed decision map that evaluates blockchain platforms based on selected and categorized blockchain features to suggest the most suitable platform. The system is flexible and responsive to changes, ensuring that if a platform becomes unavailable or gains new features, it will generate updated results accordingly.
- We introduce a decision-making map recommender system that allows us to identify the best blockchain platform(s) that are adequate for the implementation of healthcare insurance fraud detection. The decision-map is generic and can be applied to any other domain.
- We develop a taxonomy for blockchain development platforms, which we employ to determine the characteristics of the platforms that are available for implementing applications in the health insurance field. The platforms' taxonomy is based on the investigation of one hundred and two blockchain platforms in the literature.
- We exemplify the applicability of our automatic decision-map recommender system by developing and implementing blockchain smart contracts for the detection of 12 fraud scenarios.
- We evaluate the implementation of our recommender system by applying it to 42 blockchain platforms. Consequently, we develop and implement the detection of fraud on the top two recommended platforms by our decision-making map recommender system, and evaluate their performances.

- We make the recommender system toolkit and source code available on GitHub blockchain developers.

Related Works

To our knowledge, no work in the literature has automated the selection of a suitable blockchain development platform for a specific use case, such as health insurance fraud detection. In [CITATION Far20 \l 1033], the authors introduced categorizes blockchain features into four categories: 1) ‘must-have’, which indicates that the platform needed to include the specified blockchain feature to be deemed suitable, 2) ‘should-have’ implies that the defined blockchain features are highly recommended, 3) ‘could-have’ represent the optional blockchain features, and 4) ‘won't-have’, which meant to list the features which are not required by the developer. However, this method may not be precise since blockchain platforms often possess multiple features; for instance, some platforms offer various consensus mechanisms. Thus, classifying a single consensus in the ‘won’t-have’ category could unjustly disqualify a blockchain platform that might otherwise be suitable for the use case. Additionally, our system implements software that can be utilized by any clinic or hospital interested in adopting blockchain platforms. Moreover, the system is adaptive, as it allows for adding both blockchain platforms and features, as well as the modification of existing ones.

Some work introduces Machine Learning and Deep Learning models for identifying fraud and overcoming the constraints of manual detection methods. Learning models automate the detection process and enhance the analysis of patterns. As shown in Table 1,[CITATION Jia23 \l 1033] proposes a deep learning graph model, which relies on an attributed heterogeneous information network with a hierarchical attention mechanism.[CITATION Sow19 \l 1033], develops a Decision Support System (DSS) employing Genetic Support Vector Machines (GSVMs) to enhance the detection and classification of health insurance fraud in Ghana. [CITATION Set23 \l 1033] proposes an unsupervised multivariate analysis model named WMTDBC. However, the use of AI for detecting healthcare insurance fraud has raised security concerns, largely due to the sensitive client data used in training the models, consequently suffering from privacy and security issues. In addition, these works do not consider the bias introduced by the use of ML/DL algorithms. As a result, our emphasis will be on solutions that leverage smart contracts which are self-executing agreements with predefined rules that activate when conditions are fulfilled. These contracts are immutable, meaning they cannot be altered once deployed, providing a secure and privacy-preserving blockchain solution for detecting healthcare insurance fraud. Moreover, throughput, latency, and CPU and memory usage have not been taken into account in these works.

Table 1 Summary of related works on fraud detection machine learning and deep learning algorithms in health insurance claims

Algorithms under study	Number of fraud scenarios detected	Considering privacy & security	Considering bias issue	ML/DL	Throughput	Latency	CPU usage	Memory Usage	Dataset used	Metrics
MHAMFD [CITATION Jia23 \l 1033]	NR	x	x	DL	x	x	x	x	Medical-1: Balanced dataset with a ratio of positive to negative samples of 1:2.	Accuracy:0.8961 f1 score: 0.8694
									Medical-2: Unbalanced dataset with a ratio of positive to negative samples of about 1:70	f1 score: 0.8361 Recall: 0.8764 Precision: 0.9194
GSVMs [CITATION Sow19 \l 1033]	NR	x	x	ML	x	x	x	x	100-claim dataset	Accuracy: 71.43%
									300-Claim Dataset	Accuracy: 95.45%
									500-claim dataset	Accuracy: 99.18%
									750-claim dataset	Accuracy: 82.56%
									1000-claim dataset	Accuracy: 90.91%
WMTDBC[CITATION Set23 \l 1033]	NR	x	x	ML	x	x	x	x	The dataset used in the study is the claims data submitted by healthcare providers under the US Medicare CMS Part B healthcare program	Overall accuracy: Ranged from 0.857 to 0.946.

ML –

Machine Learning; DL – Deep Learning; MHAMFD – Multilevel Hierarchical Attention Mechanism for Fraud Detection; GSVMs – Genetic Support Vector Machines; WMTDBC – Weighted MultiTree Density-Based Clustering; NR - Not Reported

Therefore, researchers and developers are turning towards privacy-preserving and secure blockchain-based solutions that incorporate smart contracts for the detection of health insurance fraud. These contracts execute automatically under set conditions once deployed on the blockchain, benefiting from the platform's immutability, decentralization, and transparency, and cannot be changed after they are set up. [CITATION Mac20 \l 1033] focuses on determining whether a claim adheres to the applicable provisions of the healthcare insurance policy. [CITATION Sal20 \l 1033] proposes a solution for preventing health insurance fraud by using two fraud scenarios. [CITATION Wei19 \l 1033] uses the Ethereum blockchain to develop a framework for recording claims data and transaction patients as validators to assist in the detection of fraud. However, none of these works takes into account all possible fraud scenarios, neither the Quality-of-Services (QoS) of fraud detection in terms of throughput, and latency, nor the computing resources utilizations such as CPU and memory. In addition, the use of blockchain platforms is unjustified, and the choice of the development platform is not justified. our recommender system is adaptive to the evolving of blockchain platforms, offering a comprehensive approach. Furthermore, the smart contracts are portable and can operate across different platforms. In this paper, we implement smart contracts based on a blockchain development platform that is selected by our adaptive automatic decision-map recommender system. Based on these fraud scenarios, we implement smart contracts, for insurance fraud detection, on the top two selected blockchain development platforms by our recommender system. The strengths and weaknesses of recent works utilizing blockchain development platforms for detecting healthcare insurance fraud are summarized in Table 2.

Table 2. Summary of related works on blockchain-based healthcare insurance fraud detection

Work	Throughput	Latency	CPU utilization	Memory utilization	Number of fraud scenarios considered	Smart contract	Recommender system	Platform	Reason for choosing the platform
[19]	x	x	x	x	1	✓	x	Ethereum	NR
[20]	x	x	x	x	2	✓	x	BigchainDB	NR
[21]	x	x	x	x	3	✓	x	NR	NR
Our work	✓	✓	✓	✓	12	✓	✓	Hyperledger Fabric & Neo	Based on our proposed decision-making map recommender system tailored to healthcare insurance fraud detection

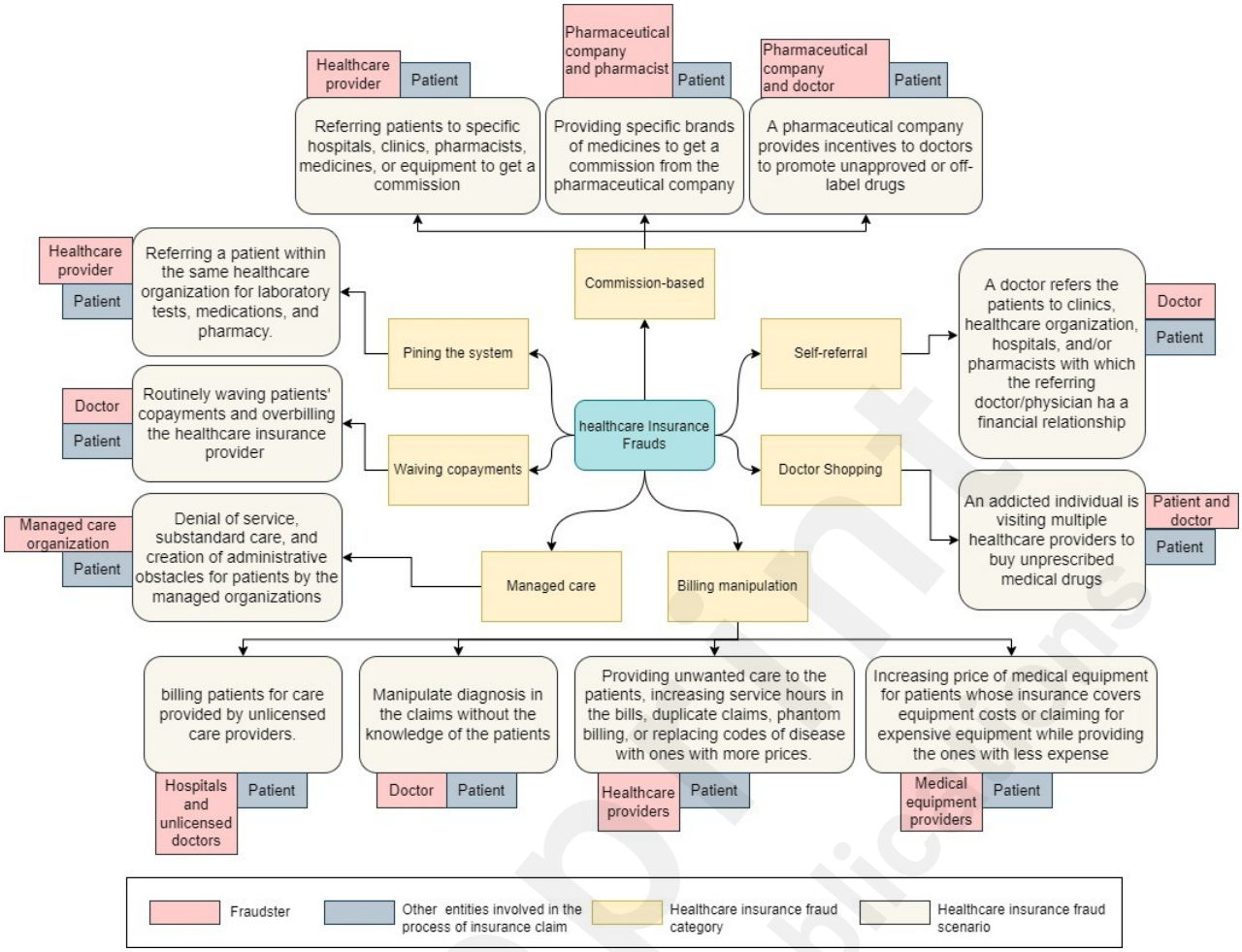
NR: Not Reported

Methods

The taxonomy of blockchain platforms is based on reviewing published research articles, and white papers that mention blockchain platforms. Our study reveals 102 blockchain platforms that we

classify according to the application domains they are developed for, such as financial services, social media, IoT, and platforms that can be utilized across several domains. In addition, we gather information on various features such as whether the platform is open-source, the consensus mechanism employed, the type of blockchain utilized, and the availability of smart contracts. For the detection of healthcare insurance fraud implementation, the fraud scenarios are based on [CITATION Lei21 \l 1033] which proposes a taxonomy of 12 fraud scenarios which are divided into 7 categories as shown in Figure 1. To begin with the first category, ‘commission-based’ which includes three fraud scenarios. The first scenario involves a healthcare provider directing patients to specific hospitals, clinics, pharmacies, medications, or equipment suppliers in return for a commission. The second fraud scenario involves pharmacies dispensing specific brands of medicines in exchange for commissions from pharmaceutical companies. The third fraud scenario involves pharmaceutical companies offering incentives to doctors to recommend drugs for unapproved or off-label uses. Next, the second category ‘Pining the System’ which involves healthcare providers guiding patients to internal entities like laboratories or pharmacies to keep profits within the organization. Following that, the third category 'Waiving Copayments' is where the doctor regularly waives patients' copayments and overcharges the healthcare provider. The fourth category, 'Managed Care' consists of organizations limiting costs by denying necessary care, providing substandard treatment, and creating administrative barriers for patients. Moreover, the fifth category 'Billing Manipulation’ consists of four fraud scenarios. The first involves unlicensed hospitals and doctors billing patients for care. The second scenario occurs when a doctor alters a diagnosis on a claim without the patient's knowledge. The third includes healthcare providers offering unnecessary care, inflating service hours, submitting duplicate claims, phantom billing, or substituting diagnosis codes for higher reimbursements. The final scenario involves medical equipment providers inflating prices for insured patients or claiming expensive equipment while supplying cheaper alternatives. Additionally, the sixth category. ‘Doctor Shopping' involves patients consulting multiple healthcare providers to obtain prescriptions for non-medical use. Finally, the seventh category 'Self-referral' occurs when doctors direct patients to clinics or healthcare facilities in which they have a financial interest, potentially leading to conflicts of interest.

Figure 1**.** Taxonomy of healthcare insurance frauds



Results

Decision-Making Map Recommender System for Selecting the Best Blockchain Platform for Healthcare Insurance Fraud Detection

The proliferation of blockchain platforms has led to a multitude of choices for developers. However, it is important to note that the various blockchain platforms available today have different features, capabilities, and use cases[CITATION Ism192 \l 1033],[CITATION Rim23 \l 1033]. Therefore, developers need to evaluate the available options and select the platform that best fits their specific needs. In this section, we provide an overview of our taxonomy, which encompasses 102 blockchain platforms. Subsequently, we present our features-based decision map recommender system to select the best platform.

Taxonomy of Blockchain Platforms

In 2008, Bitcoin [CITATION Sat08 \l 1033] made its debut, and the subsequent addition of smart contract technology by Ethereum [CITATION Eth \l 1033] contributed significantly to the rapid growth and development of blockchain technology. As a result, more than 100 distinct blockchain platforms were developed for various purposes. To provide a comprehensive understanding of these platforms, we present a taxonomy of 102 blockchain platforms which we organize based on their respective application domains. Along with the application domain, our classification also takes into account the open-source nature of the platform, the consensus mechanism utilized, the type of blockchain, and the platform's capability to support smart contract development. The taxonomy of blockchain platforms is presented as a graph in the following figures. Figure 2 presents an overview of the different generic blockchain platforms that can be utilized to build a wide range of

applications.

Figure 2. Graphical representation of the taxonomy of blockchain platforms. Generic blockchain platforms

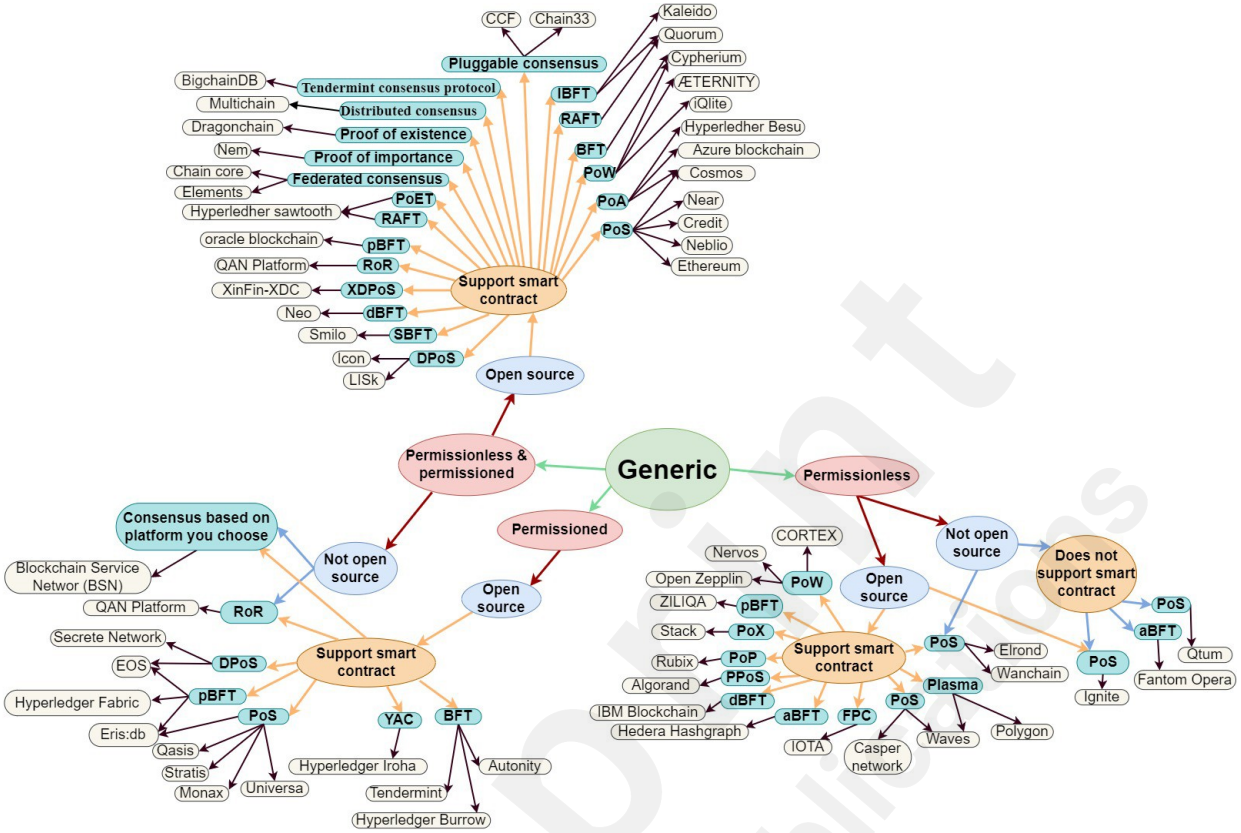


Figure 3 presents the blockchain platforms that have been specifically designed for financial services, while Figure 4 presents the platforms that are tailored to meet the needs of a particular application domain. These specialized platforms offer specialized features and functionality to cater to the specific needs of their respective industries or sectors, thus providing a more specialized and customized solution for these specific use cases.

Figure 3. Graphical representation of the taxonomy of blockchain platforms. Blockchain platforms dedicated to financial services.

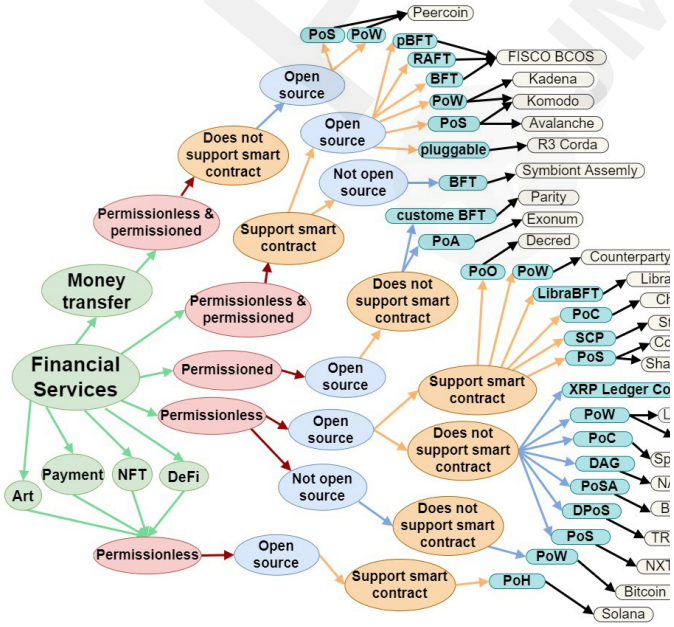
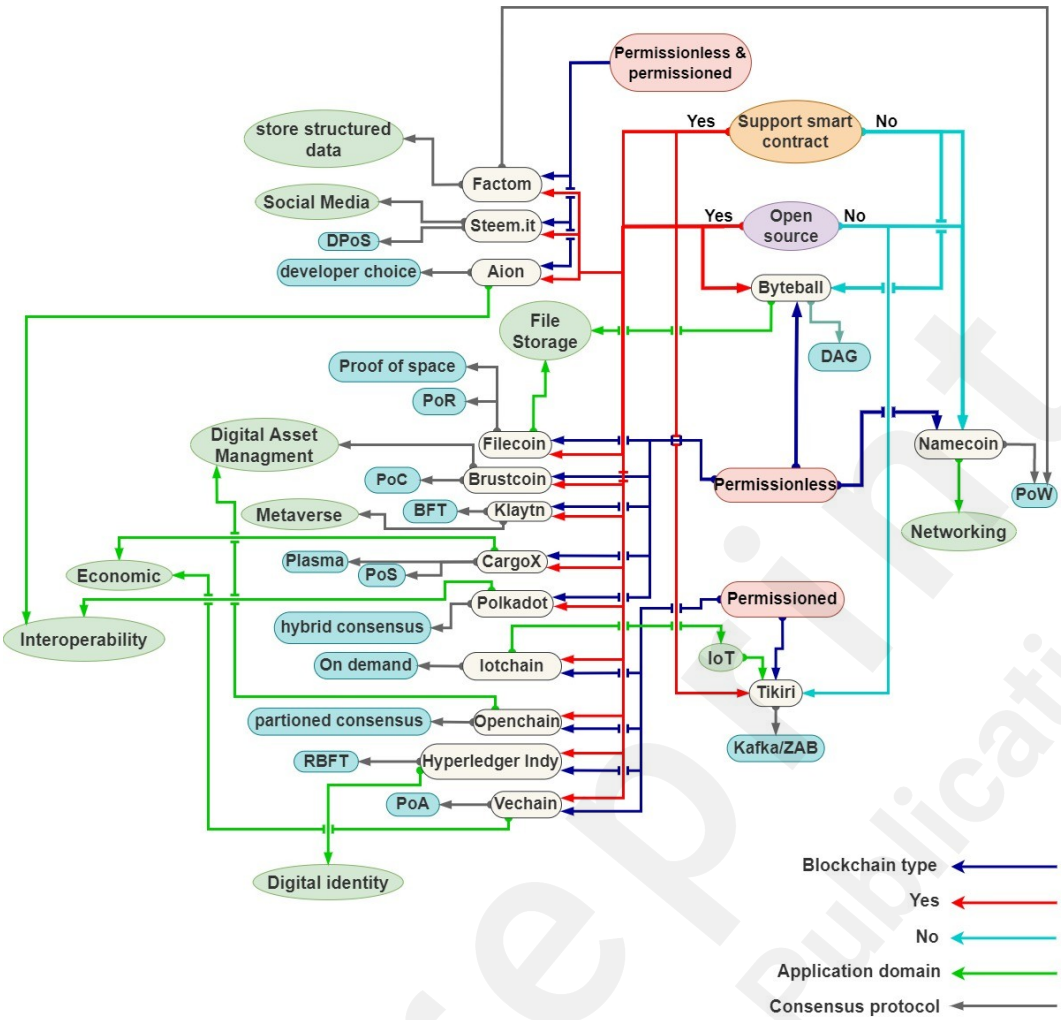


Figure 4. Graphical representation of the taxonomy of blockchain platforms. Blockchain platforms

mapped to specific application domains.



Decision-Making Map-Based Recommender System

While a blockchain platform selection method was proposed in [CITATION Far20 \l 1033], it included unnecessary categories of features and did not specifically focus on the detection of insurance fraud. Therefore, we propose a decision-making map that is tailored specifically to healthcare insurance fraud detection solutions. It classifies blockchain features into three main categories: "compulsory features", which are essential to the platform; "mandatory features", which are sufficient; and "possible features", which are desirable but not necessary. These categories differ in weight, which determines the value of one feature over another. As shown in Figure 5, our map offers a targeted approach to selecting a blockchain platform for developing health insurance fraud detection mechanisms.

In the healthcare insurance domain, privacy is a crucial aspect as insurance companies deal with sensitive patient data [CITATION Lei20 \l 1033]. Several research works in healthcare implement blockchain technology to ensure integrity, accountability, and non-repudiation in the claim process [CITATION Lei21 \l 1033], [CITATION Mac20 \l 1033]. [CITATION Lei21 \l 1033] proposes a blockchain system for healthcare insurance anti-fraud that ensures trusted medical process information entry and reading, as well as a data privacy protection scheme. In [CITATION Mac20 \l 1033], a blockchain system is proposed and implemented to prevent counterfeiting in healthcare

insurance, providing a secure, and private system.
To implement healthcare insurance fraud detection using blockchain, we should select features that ensure privacy, such as on-chain transactions, and permissioned platforms. This is in addition to other technical features that should be available in the platform, such as the smart contracts and user interface development tools features. In summary, we determine the most suitable features, both in terms of their relevance to the task of healthcare insurance fraud detection and the technical capabilities of the platforms. We divide these features into three categories which are compulsory, mandatory, and possible features:
Compulsory features:

- Application layer: this capability enables the creation of a user interface and the execution of smart contracts for healthcare insurance.
- Network layer: enables the establishment of a peer-to-peer decentralized network.
- Protocol layer: enables the selection of a consensus; we are using Byzantine-based consensus because it prevents the case of a failing node and malicious node [CITATION Ism191 \l 1033].
- Interoperability technologies: technologies such as Oracle that facilitate the integration of data from off-chain resources into smart contracts.
- On-chain transaction: the transaction is conducted on the main blockchain for increased security, decentralization, and transparency.
- Permissioned blockchain: this type of blockchain limits access to the ledger to a select group of trusted nodes.
- Smart contracts: enable the development of algorithms that can identify healthcare insurance fraud.

Mandatory features:

- Enterprise system interrogation: provides easy access to data, seamless data flow, and time and cost savings.
- Private: this type of blockchain network is only accessible to authenticated users.
- Turing complete: the virtual machine of the blockchain platform is capable of solving any computational problem.
- JavaScript, Python, and Solidity: these languages are specifically mentioned because they are intuitive and easily learned by programmers.

Possible features:

- Java and Golang: these languages, such as the last three mentioned in the "mandatory features" list, are intuitive and easily learned by programmers.
- Virtual machine: it is used to execute smart contracts.
- Privacy technology: ensures data privacy and certifies the eligibility of peers to participate in the network, particularly when handling sensitive patient data.
- Zero-knowledge proof: this encryption scheme allows one party (the prover) to assure another party (the verifier) that they know a certain value (X) without revealing the value itself.
- Cryptographic token: these tokens have the potential to be used as a means of payment.
- Cross-chain interoperable: this feature enables the connection of two separate blockchains to facilitate information exchange.

Based on the selected features above (compulsory, mandatory, and possible), our enforced decision-making map selects 42 platforms out of 102. This extraction of 42 platforms is derived from our proposed taxonomy of blockchain platforms. This taxonomy maps the blockchain platforms into their corresponding applications' domains and blockchain features. To ensure the privacy and

security of patient files, the decision-map recommender system selects the blockchain platforms that meet these specific criteria. Therefore, the selection process excludes platforms that are based on permissionless blockchain type, which is open to the public and may compromise data confidentiality. Instead, the recommender system prioritizes platforms that are suited for generic application domains and financial services, as per our taxonomy. In addition, the recommender system focuses on platforms that support the development of smart contracts.

After identifying the relevant blockchain features for healthcare insurance fraud detection, the recommender system initiates a mapping process to match each feature with the platforms that support it. Figure 5 illustrates the outcomes of this mapping. Initially, after organizing the blockchain features into categories, the recommender system proceeds to map each feature with its corresponding functionality. Next, the recommender system maps the features to the blockchain platforms. Based on that, it determines the suitability of each platform. Only the platforms that have all of the compulsory features are considered suitable. As shown in Figure 5, R3 Corda and BigchainDB were eliminated from consideration due to their lack of some of the compulsory features. Our mapping process reveals that Hyperledger Fabric [CITATION Hyp \l 1033] is the most optimal platform, followed by Neo [CITATION NEO \l 1033], Xin-Fin-XDC [CITATION Xin21 \l 1033], Quorum [CITATION Con \l 1033], and Ethereum. These results demonstrate the effectiveness of our mapping process in identifying the ideal blockchain platform for this specific use case.

Table 3 streamlines the mapping process for healthcare insurance fraud detection by listing the top 5 platforms and highlighting the selected features. The table is designed to simplify the decision-making map by providing a concise and easy-to-read format for comparing the features.

Figure 5. Proposed decision-making map for selecting a platform for healthcare insurance fraud detection

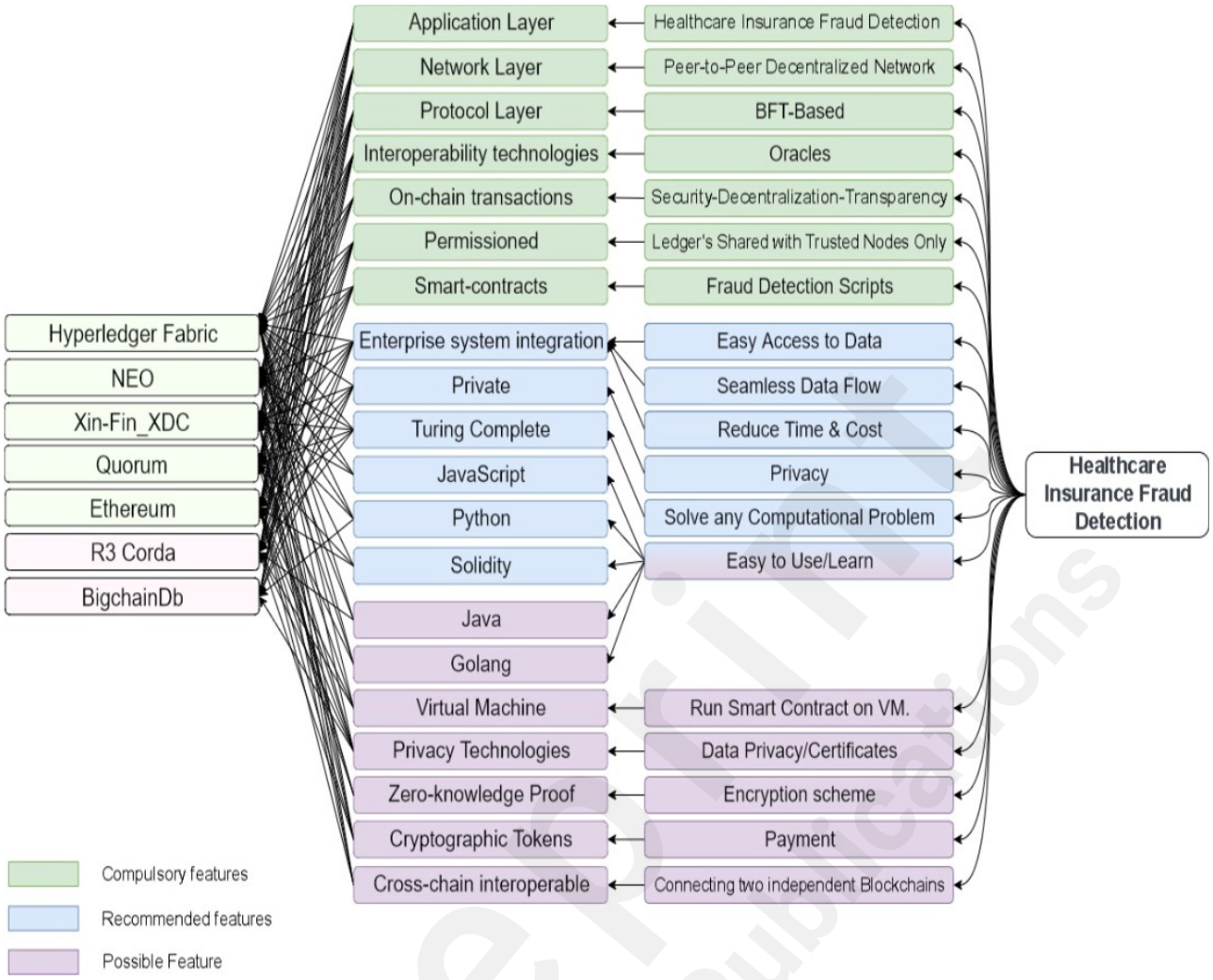


Table 3. Decision-making map results simplified

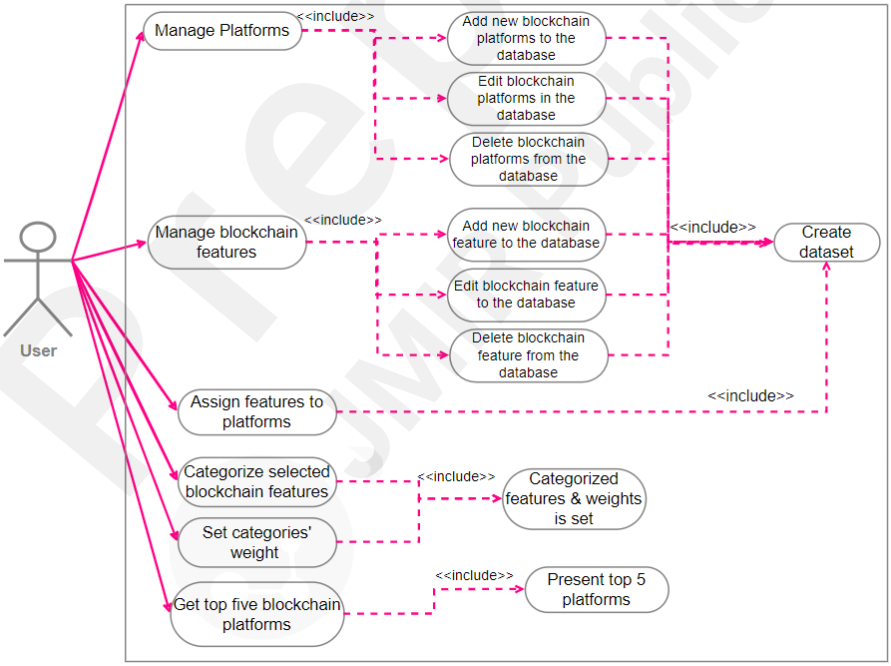
Category	Feature name	Fabric Hyperledger	NEO	Ethereum	Quorum	Xin-Fin-XDC
Compulsory features						
	Application layer	✓	✓	✓	✓	✓
	Interoperability technology	✓	✓	✓	✓	✓
	Network layer	✓	✓	✓	✓	✓
	On-chain transaction	✓	✓	✓	✓	✓
	Permissioned	✓	✓	✓	✓	✓
	Protocol layer	✓	✓	✓	✓	✓
	Smart contract	✓	✓	✓	✓	✓
Mandatory features						
	Enterprise system integration	✓	✓	✓	✓	✓
	JavaScript	✓	✓	X	X	✓
	Private	✓	✓	✓	✓	✓
	Python	✓	✓	X	X	✓
	Solidity	✓	X	✓	✓	X
	Turning complete	✓	✓	✓	✓	✓

Category	Feature name	Fabric Hyperledger	NEO	Ethereum	Quorum	Xin-Fin-XDC
Possible features						
	Zero-knowledge proof	✓	X	✓	✓	X
	Virtual machine	✓	✓	✓	✓	✓
	Java	✓	✓	X	X	X
	GoLang	✓	✓	X	✓	X
	Cryptographic token	✓	✓	✓	X	✓
	Cross-chain interoperable	✓	✓	✓	X	✓
	Privacy technology	✓	✓	✓	✓	✓

Use Case Diagram for the recommender system

Figure 6 illustrates the use case diagram of our recommender system. Users can perform actions such as adding, editing, and deleting blockchain platforms and features. Following that, they are required to select their desired features, categorize them, and assign weights to mandatory and possible features. Ultimately, users will receive the outcome of the most suitable blockchain platform for their specific use case based on the chosen features.

Figure 6. Recommender system use case diagram



Decision-Making Map Recommender System Implementation

In this subsection, we present our implementation of the decision-map recommender system, which is a desktop software solution that provides a streamlined and efficient method to select the most suitable blockchain platform for a specific use case. Our software utilizes WinForms C# technology and SQL as the database, to deliver a user-friendly experience and recommend the top blockchain platforms. To demonstrate the effectiveness of our software, we utilized it to identify the top 5 blockchain platforms that are most suitable for healthcare insurance fraud detection.

defines each function. As previously discussed, the blockchain feature selection process involves

categorizing features into three categories: compulsory, mandatory, and possible. Compulsory features are those that must be present in the blockchain platform for it to be considered. These features are typically critical to the platform's functionality. Mandatory features, on the other hand, are those that are essential for a specific use case or application. They are not necessarily required for the platform to function, but they are necessary for the platform to be suitable for a particular purpose. Finally, possible features are those that provide additional functionality or value to the platform. They are not necessary for the platform to function, but they can enhance its performance or provide additional benefits.

Table 4. Decision-map recommender system functions and their definitions

Function	Definition
Create the dataset	Blockchain platform and blockchain feature names are initially entered. Subsequently, the platforms are associated with their corresponding features
Select features and their categories, and set weights	Specify the category of the blockchain feature by selecting one of the three options, namely compulsory, mandatory, or possible. Then, assign the selected feature to the designated category. Additionally, assign weights to the mandatory and possible features
Get top platforms	Retrieve blockchain platforms with 'compulsory features' and count the number of mandatory and possible features found for each platform. Calculate a score for each platform based on its features and weights, and add it to an array. Sort the array based on the calculated score to display the top-performing blockchain platforms

Creating the Dataset

This section presents screenshots to illustrate the dataset creation process, including user interactions with the recommender system. The blue annotations represent instructions. The pink annotations indicate the text boxes for input, buttons for actions, and grid controls for displaying the added platforms and features. In this step, we focus on adding the necessary platforms and blockchain features to build a comprehensive dataset. Figure 7 demonstrates the user's process of entering the platform name and selecting "Add Platform" to populate a table displaying the added platforms. The same procedure applies to adding blockchain features, resulting in a comprehensive table showcasing both platforms and features. After that, users are provided with the capability to edit platform names or delete them, as well as modify the names or choose to delete blockchain features. As shown in Figure 8, by double-clicking on a platform and single-clicking on a feature, users can select and make changes to the respective names according to their preferences. After that, users should establish the association between each blockchain platform and its corresponding blockchain features. They can begin by selecting a platform by double-clicking on the row corresponding to the platform name and subsequently choosing the blockchain features that apply to that particular platform, and this is by double-clicking on the rows that correspond to the blockchain features that should be mapped to the selected blockchain platform (Figure 9). Subsequently, a table will be populated with the IDs of the selected blockchain platform, the chosen blockchain feature, and the name of the blockchain feature.

Figure 7 Dataset creation - step 1: adding platforms and blockchain features that will be used to create a dataset

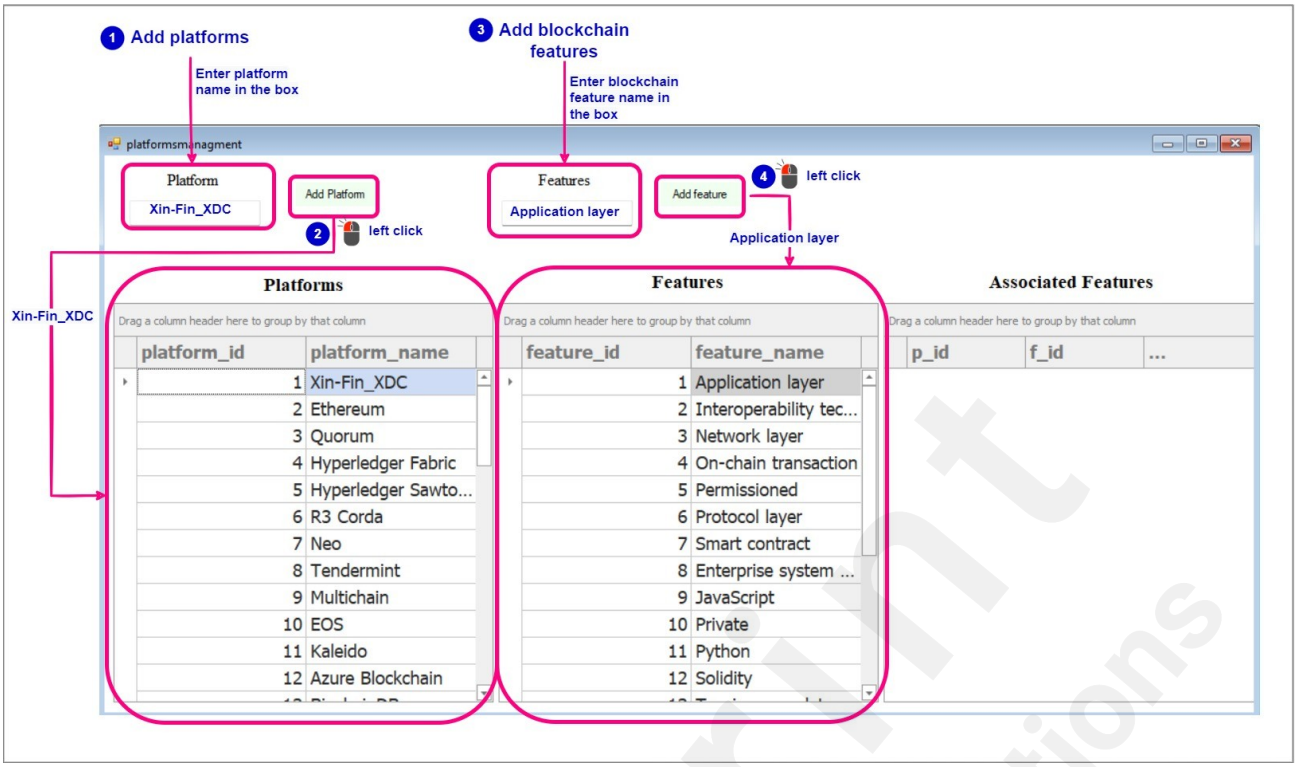


Figure 8 Dataset creation - modifying and removing platforms and features from the dataset.

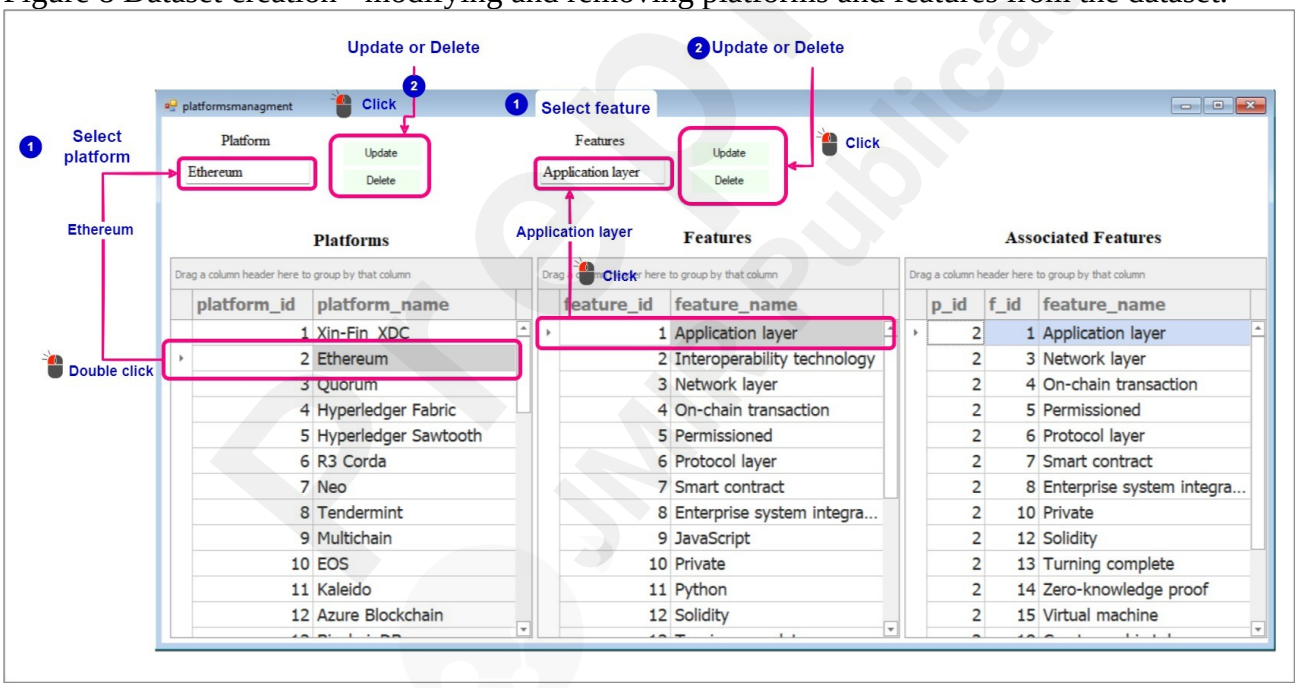
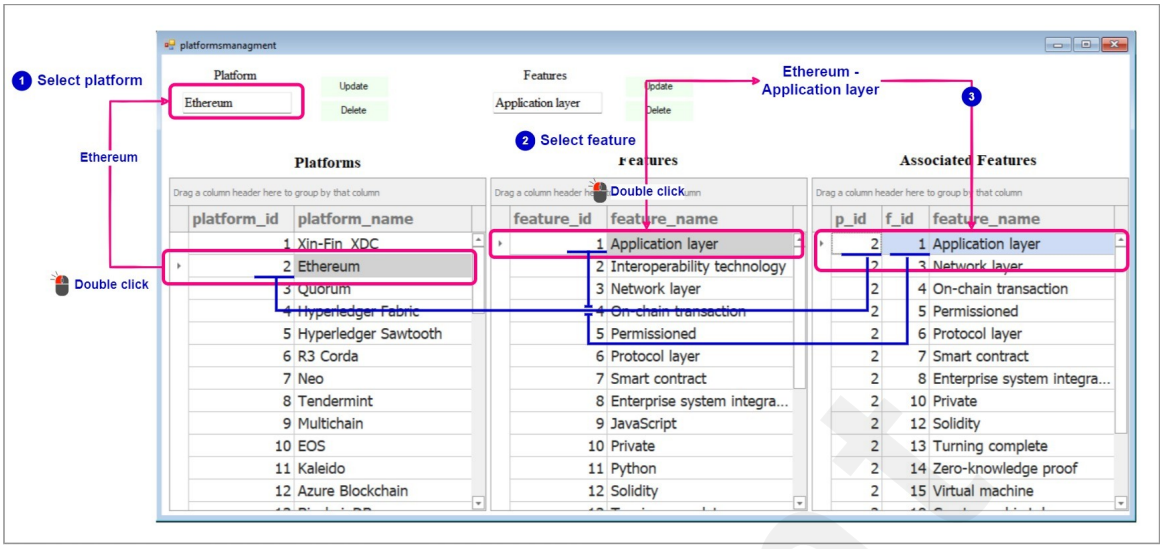


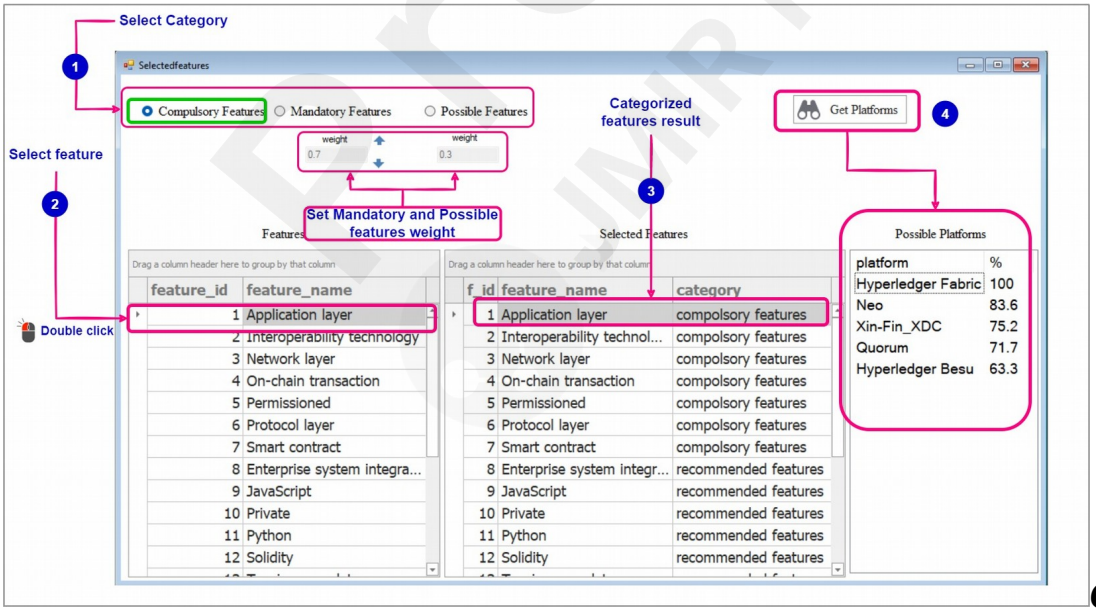
Figure 9. Dataset creation - linking specific features with respective blockchain platforms



Select Features and Their Categories, And Set Weights

Figure 10 displays a screenshot illustrating user interaction during the process of selecting features, assigning them to their respective categories, and assigning weights to those categories. In the initial step, users will choose a category, followed by selecting the desired feature to be assigned to that category. This selection process involves double-clicking on the feature name in the table. Subsequently, a table will display the categorized features, providing a clear overview of the features that have been assigned to their respective categories. Once the categorization of features is completed, users can proceed to set the weights for the mandatory and possible feature categories. Afterward, by clicking on the "Get Platforms" button, users can view the resulting platforms based on the assigned weights and feature categorization.

Figure 10. Allowing specification of preferred blockchain features, followed by categorization and weight assignment.



Get Top Platforms

Figure 11 (Part 1 and Part 2) shows the sequence diagram to obtain the suitability percentage of each possible blockchain platform. The initial step involves creating a list of blockchain platforms that meet the requirements of the compulsory features. Once that is done, we determine the total number

of mandatory and possible features that we have chosen (Figure 11 Part 1).

After that, we iterate through the list, and for each platform, we calculate the number of mandatory and possible features (Figure 11 part 2). Finally, using Equation 1 we calculate the suitability percentage of each platform (ρ). The first part of the formula calculates the contribution of the mandatory features to the suitability percentage. It takes the number of mandatory features found for the platform (M^{found}), multiplies it by 100 to convert it to a percentage, and then divides it by the number of mandatory features selected (M^{total}) multiplied by the weight assigned to mandatory features (ω^{M}), which is 0.7.

The second part of the formula calculates the contribution of the possible features to the suitability percentage. It takes the number of possible features found for the platform (P^{found}), multiplies it by 100 to convert it to a percentage, and then divides it by the number of possible features selected (P^{total}) multiplied by the weight assigned to possible features (ω^{P}), which is 0.3.

By combining these two contributions, the suitability percentage provides an overall assessment of how well a blockchain platform meets the selected features, with a higher percentage indicating a better match.

Equation 1. Blockchain platform suitability percentage equation

$$\rho = \left(\frac{M^{found} \times 100}{M^{total}} \times \omega^{R} \right) + \left(\frac{P^{found} \times 100}{P^{total}} \times \omega^{P} \right)$$

ρ: possible blockchain platform

M^{found}: number of mandatory features found for the possible blockchain platform

M^{total}: number of mandatory features selected

ω^{M}: mandatory features weight

P^{found}: number of possible features found for the possible blockchain platform

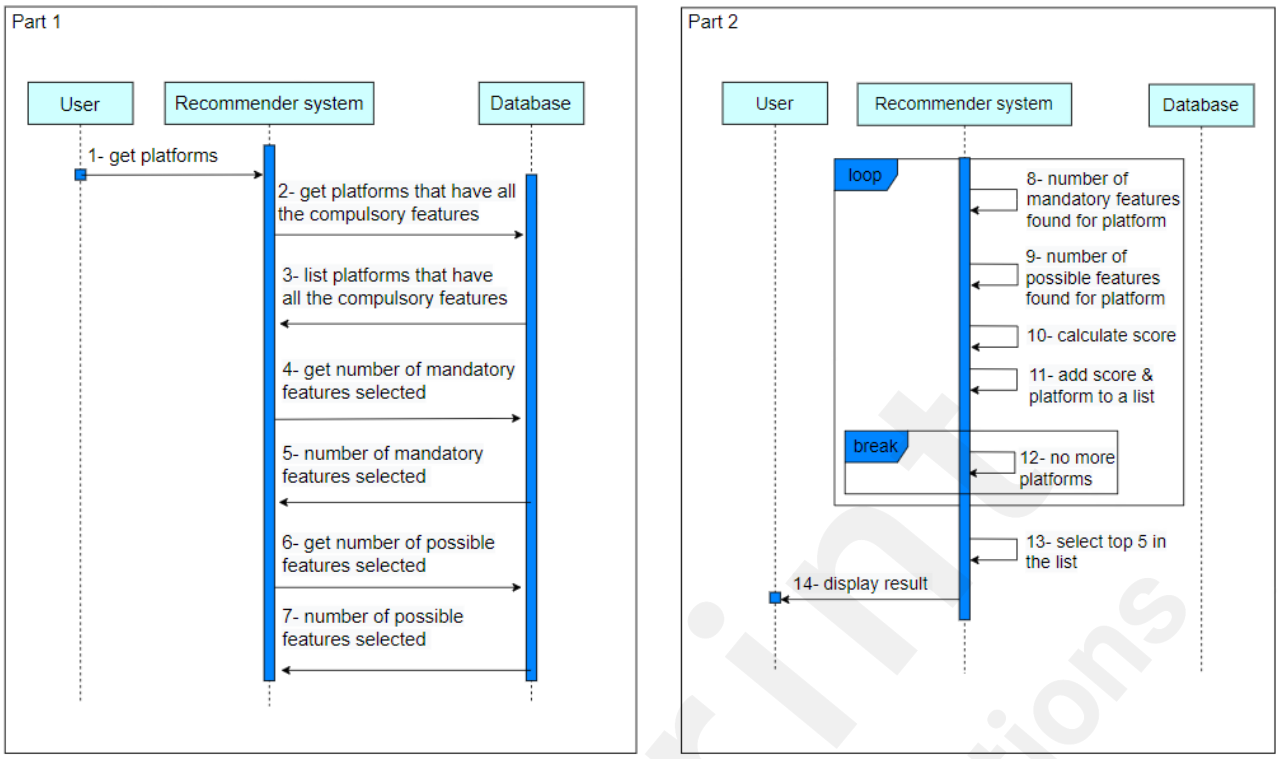
P^{total}: number of possible features selected

ω^{P}: possible features weight

ω^{c}: compulsory features weight

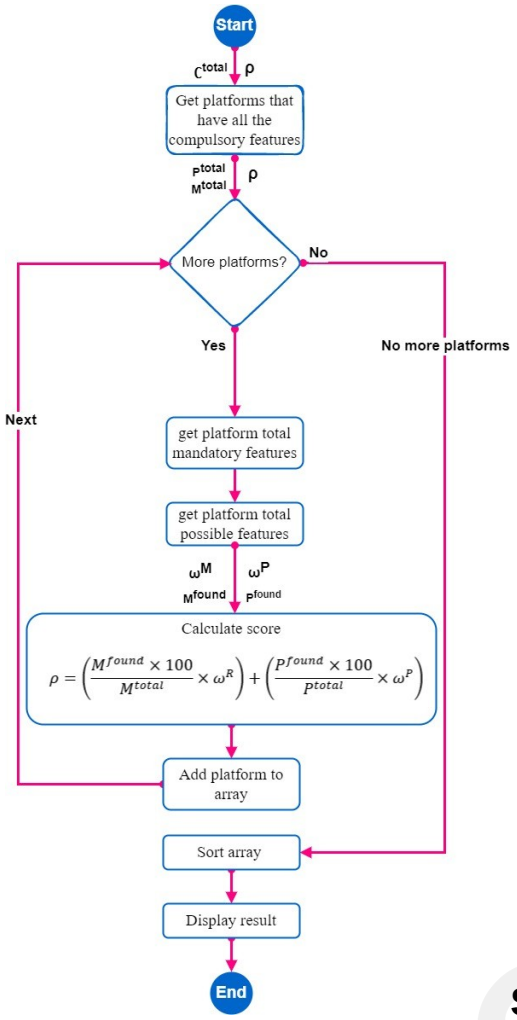
Once we have calculated the suitability percentage (ρ) for each platform, we sort the list of platforms in descending order based on their scores. Figure 12 shows the flowchart of the recommender system's algorithm, which consists of the different functions involved along with their corresponding input and output parameters.

Figure 11. The diagram to get the fitness of the possible blockchain platforms for the use case



The platform with the highest score will be at the top of the list, while the one with the lowest score will be at the bottom. Finally, we display the top five platforms in the list, which are the ones that have the highest scores and are therefore most suitable based on the selected features.

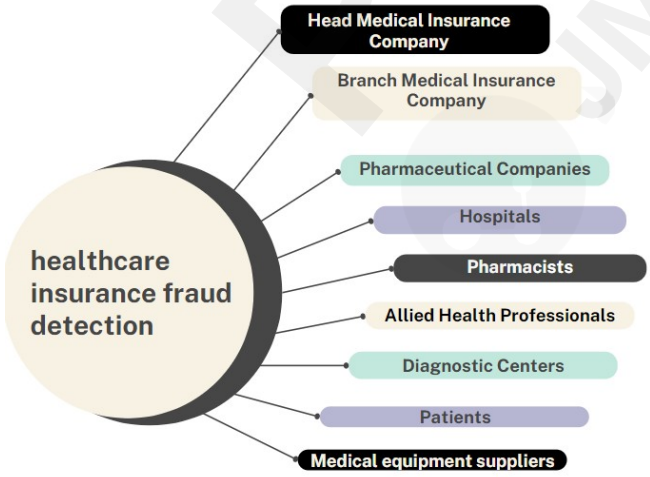
Figure 12 The flowchart to get the fitness of the possible blockchain platforms for the use case



Smart Contracts for Healthcare Insurance Fraud Detection

The authors of [CITATION Lei21 \l 1033] identified the fraud scenarios used for detecting healthcare insurance fraud, as depicted in Figure 1. The network for detecting healthcare insurance fraud is made up of nine participants, as illustrated in Figure 13.

Figure 13. Participants in the healthcare fraud detection network



Algorithms for Fraud Scenarios

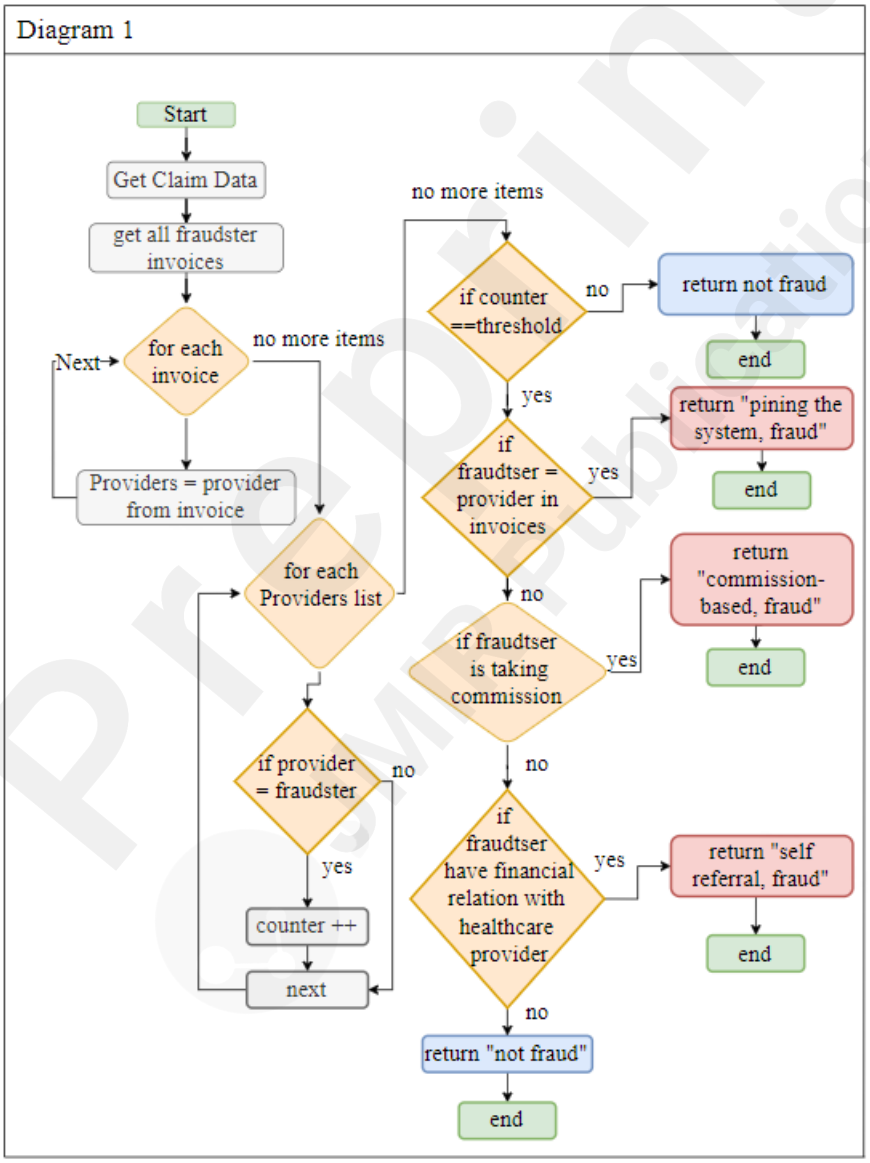
For certain fraud scenarios, we need to discover a detectable pattern, while in others, data from off-chain sources may be required. The required data for processing claims consists of detailed records

of patient visits, including the dates they occurred, the departments involved, the services rendered, and patients' information. Consequently, they are on-chain. However, documentation of billed services, detailed service invoices, and pharmacy records are off-chain in the database.

Three Referring Fraud Scenarios

In this diagram, as shown in Figure 14 (diagram 1), we utilize an algorithm to recognize three fraud scenarios that have the same pattern - the referring part. We then check for the first scenario in which the fraudster is referring patients within the same healthcare organization. If this is confirmed, fraud is pinning the system. If not, we investigate whether a financial relationship exists between the fraudster and the other organizations. If such a relationship is detected, it is self-referral fraud. If no financial relationship is found, we investigate whether the fraudster received a commission from the organization; if so, it is a commission-based fraud.

Figure 14. (Diagram 1) Three fraud scenarios are related to referring.

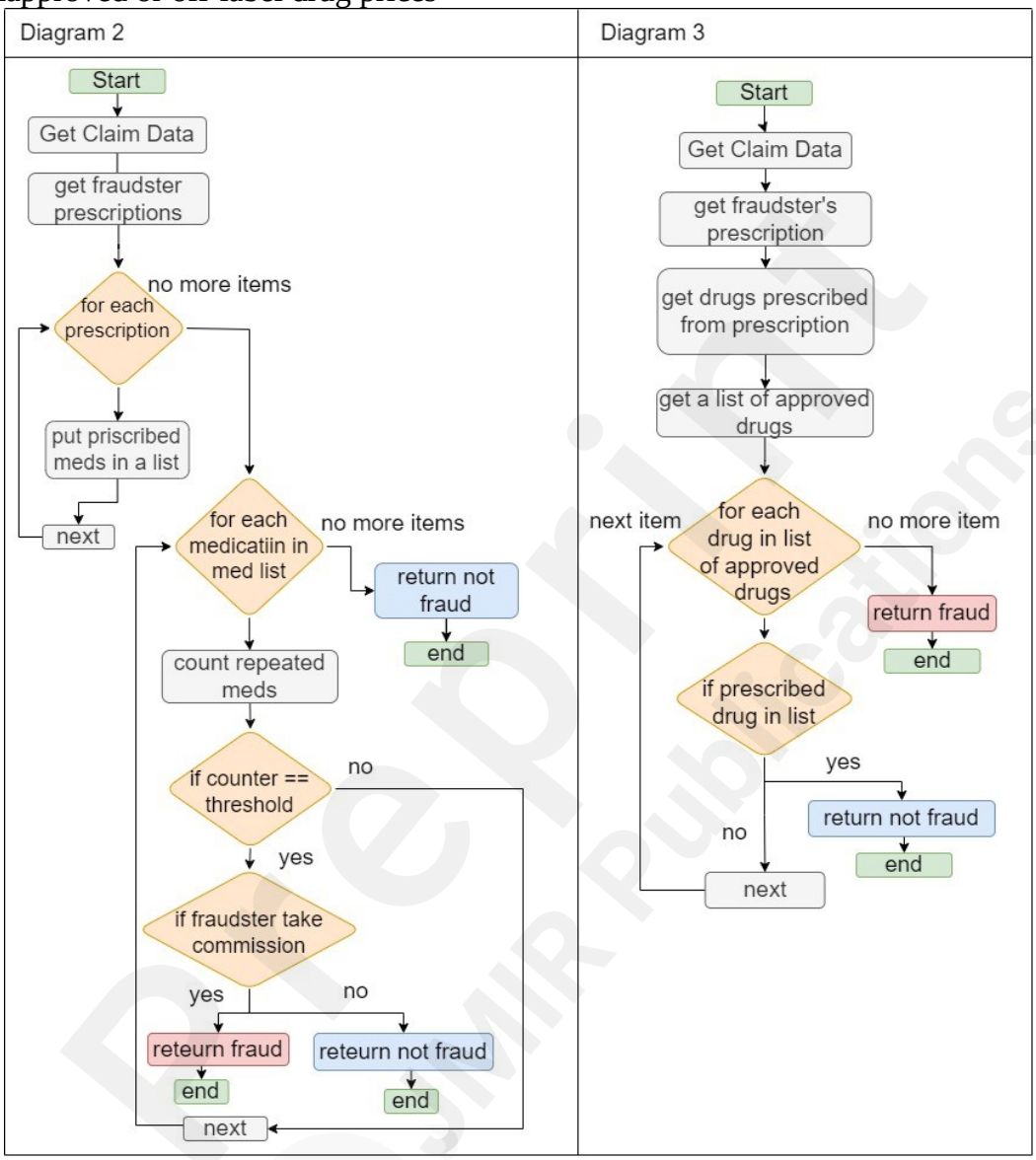


Commission Based

In this fraud scenario (Figure 15 (diagram 2)), we obtain all of the medication that the possible fraudster has prescribed, we then check if a specific medication is prescribed more frequently, and determine if the fraudster is receiving a commission, if that's the case, it is a fraud. In this diagram (Figure 15 (diagram 3)), we should obtain a list from the minister of health containing the approved

and labeled drugs, then compare it to the ones prescribed by the fraudster; if we find a drug that does not exist on the list, we have a fraud.

Figure 15. (Diagram 2) providing specific brands of medicines to get a commission from the pharmaceutical company, (Diagram 3) a pharmaceutical company provides incentives to doctors to promote unapproved or off-label drug prices

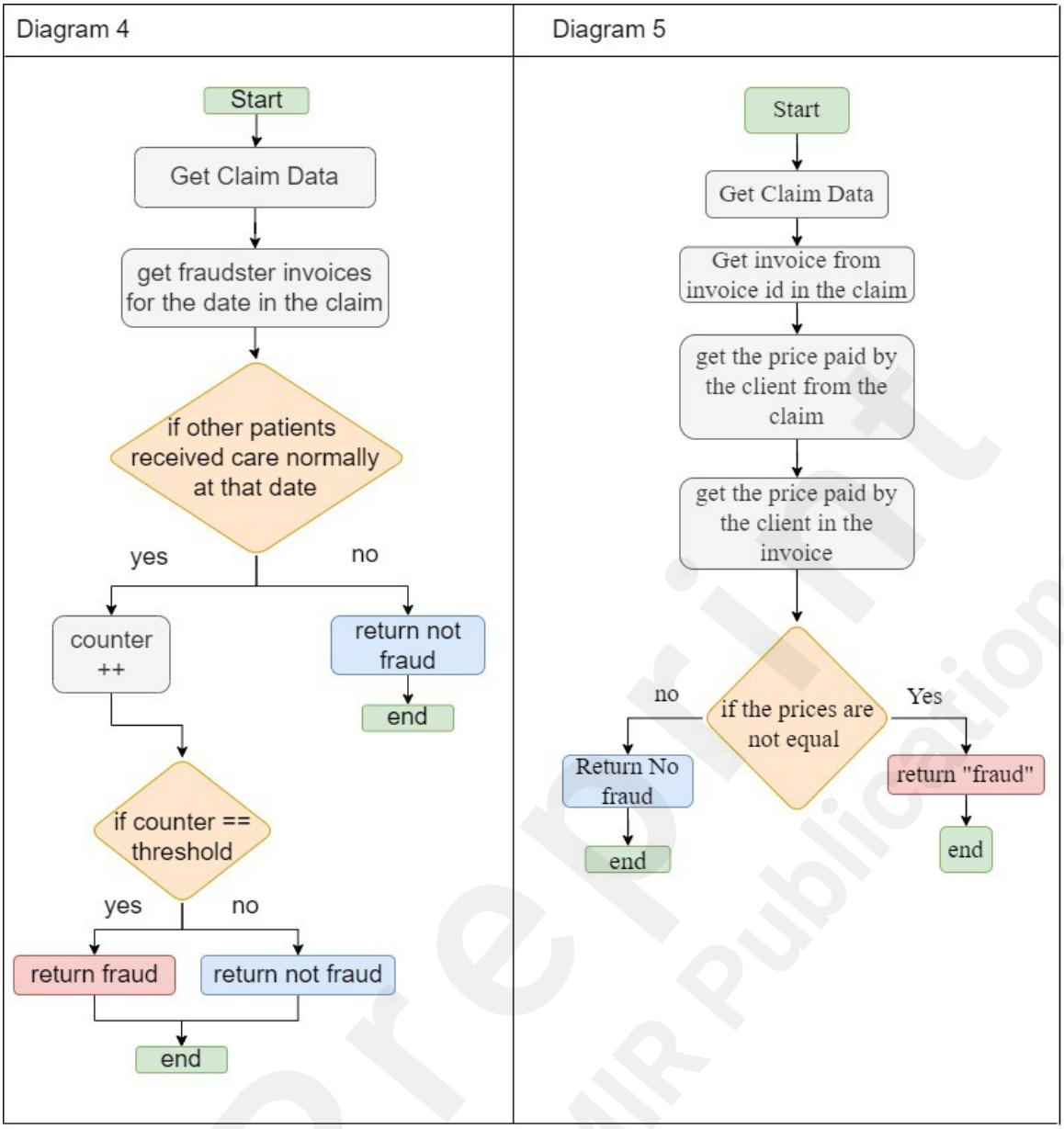


Managed Care and Waiving Copayment

In Figure 16 (diagram 4), we investigate whether other patients on that date received the same service that the patient requested, and if the number of patients reaches a certain threshold, we will be able to demonstrate that the managed care scenario is occurring.

The code for detecting waiving copayment fraud in Figure 16 (diagram 5) involves comparing the price listed in the claim with the price mentioned in the corresponding invoice. If there is a mismatch between the two prices, it indicates a potential instance of waiving co-payment fraud. The code performs a comparison operation to check if the claim price and the invoice price are equal. If they are not, it raises an alert or triggers further actions to investigate the possibility of fraud.

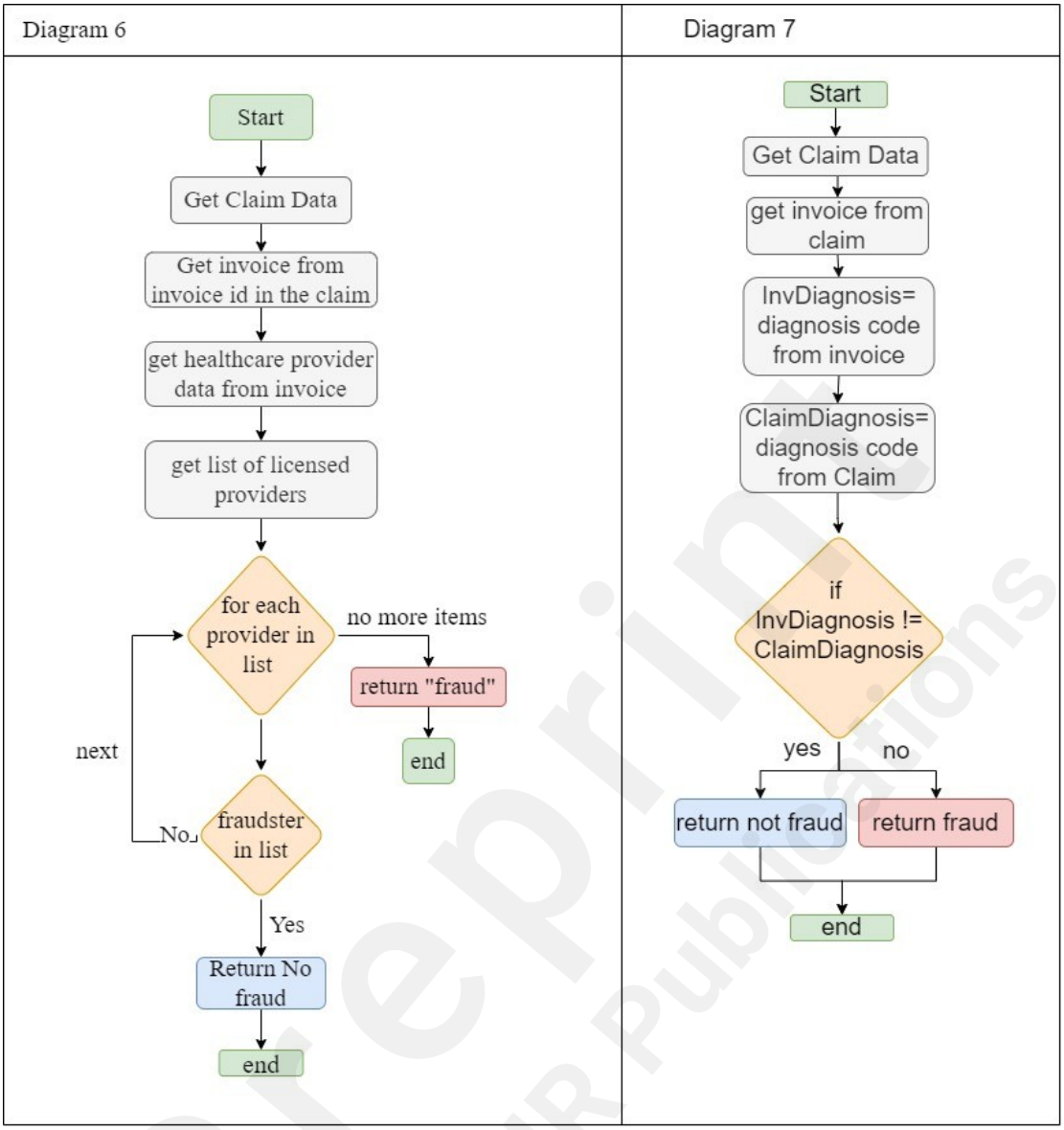
Figure 16. (Diagram 4) managed care, (Diagram 5) waiving copayment



Billing Manipulation

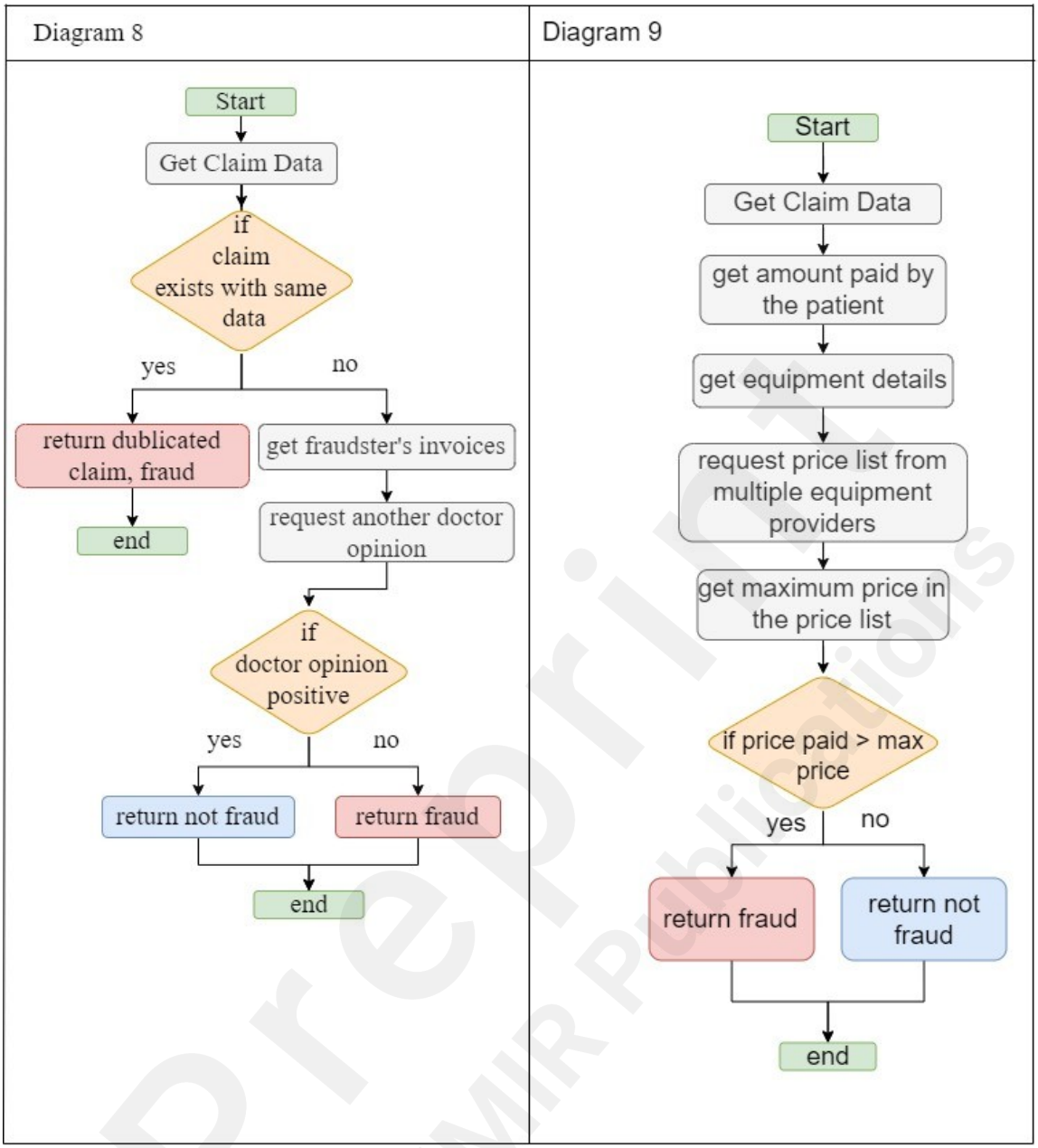
Figure 17 (diagram 6), we need a list of licensed healthcare providers to determine whether the suspected fraudster is listed. If not, we have a fraud case. In Figure 17 (Diagram 7), we compare the diagnosis code on the claim to the one on the patient files; if they do not match, we will assume fraud.

Figure 17**.** (Diagram 6) billing patients for care provided by an unlicensed care provider. (Diagram 7) manipulation of diagnosis in the claims without the knowledge of the patients.



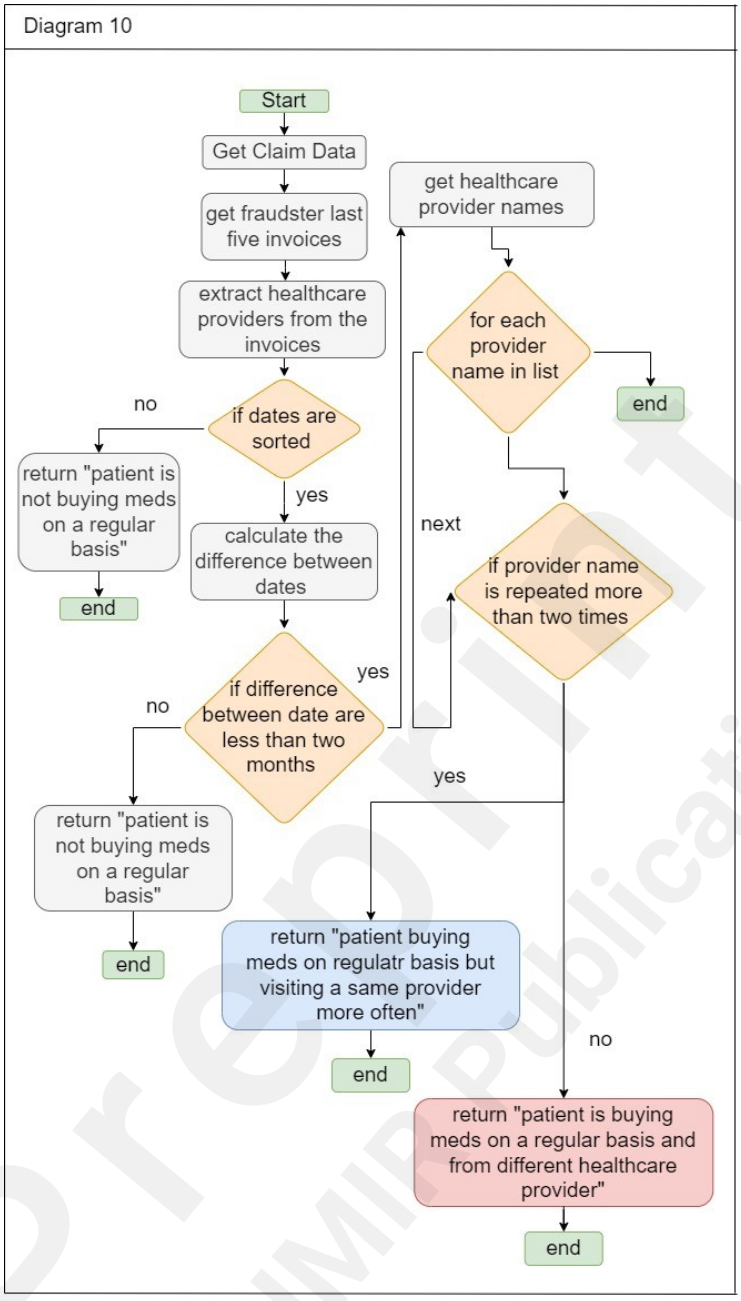
In this fraud scenario (Figure 18 (diagram 8)), we may require the opinion of another doctor, so after determining whether or not the claim has been replicated, we gather all of the necessary data to be reviewed by another doctor, and based on the doctor's replay, we determine whether or not the claim is fraudulent. In Figure 18 (diagram 9), we should obtain a price list from other equipment suppliers, then compare it to the price paid by the patient; if it is higher than the price on the price list, the patient paid more than necessary for the equipment, hence it’s a fraud.

Figure 18. (Diagram 8) providing unwanted care to the patients, increasing service hours in the bill, duplicating claims, phantom billing, or replacing codes of diseases with ones with higher prices. (Diagram 9) billing manipulation in equipment prices



Doctor Shopping

Figure 19 (Diagram 10) illustrates doctor shopping fraud, in which an addicted individual visits multiple healthcare providers to obtain unprescribed drugs. To detect this fraud, we must examine five invoices. We check if the patient is visiting the provider regularly based on the dates from the invoices, and if the visits are not to the same provider. If this is confirmed, it is a case of fraud.
Figure 19.(Diagram 10) doctor shopping



Discussion

Experiments

We create algorithms for smart contracts that address the fraudulent situations mentioned in[CITATION Lei21 \l 1033]. We utilize the two platforms that were selected as the top two options according to our decision-making recommender system. Each transaction contains a single claim record and requires supporting files such as invoices for verification. We assess the performance of the platforms based on several metrics, including throughput, latency, CPU usage, and memory usage. Transaction throughput reflects the blockchain network's efficiency in terms of the number of transactions processed per second. A transaction is considered successful once it has been included in a block and committed to the ledger. Transaction latency measures the time it takes to send a transaction request and receive a transaction response, indicating the network's responsiveness. CPU and memory usage are essential for determining infrastructure requirements and maintaining reliable performance under varying loads[CITATION Mar22 \l 1033]. It is essential to ensure that a platform maintains high throughput, low latency, and minimal CPU and memory usage [CITATION

Lei08 \l 1033]. This means that QoS is maintained, and consequently, fraud is detected more quickly, costs are reduced, and, in some cases, patient lives are saved[CITATION Raj23 \l 1033]. To evaluate their effectiveness, we establish two testing scenarios. The first scenario involves peers sending a consistent number of transactions (TX) over a period ranging from 30 seconds to 120 seconds. In the second scenario, we progressively increase the number of transactions transmitted over the network from 1000 TX to 10,000 TX.

Experimental environment

Our experimental setup involves using Ubuntu 20.04 WSL2 on Windows 11 operating systems. In the case of Fabric, we employ Docker to run the platform, and all peers are connected to a single channel representing an insurance company. The Fabric version we use is v2.2, with 4 organizations and one orderer. A batch size of 500 is set for processing transactions. We use Golang as the programming language for developing smart contracts.

For the Neo private blockchain, we utilize the N3 Neo Visual DevTracker extension on Visual Studio Code. The programming language recommended in the Neo documentation, which is C#, is used. Similar to Fabric, a batch size of 500 is employed. Both platforms utilize LevelDB as the key-value data storage.

In terms of hardware, we employ a system with 16 GB of RAM and an 11th-generation Intel Core i7 processor running at 2.80 GHz. To conduct the benchmarking, we utilize Hyperledger Caliper [CITATION Hyp4 \l 1033] for Fabric and Neo-bench [CITATION Neo \l 1033] for Neo.

Security and Privacy Concerns

Healthcare data comprises sensitive patient information, making its security and privacy crucial. These data are typically stored in local or cloud databases. However, in such architecture, data faces several issues regarding data access and is subject to be deleted or modified, and is subject to cyber security attacks [CITATION Lei211 \l 1033].

Using a blockchain-based approach addresses these concerns. Blockchain stores data in an immutable ledger where data can only be added after reaching a consensus and cannot be altered or modified. Changing a block or transaction impacts all subsequent blocks, and revalidating all subsequent blocks requires enormous computational power, making it nearly impossible for a malicious node. Furthermore, access control rights in blockchain can be defined in a smart contract, ensuring trustless and secure data access for network participants. Table 5 provides a concise overview of the challenges encountered in local on premises/cloud database systems and demonstrates how blockchain technology addresses these issues.

Table 5 Comparison of Issues in Local on Premises/Cloud Databases and Blockchain for Managing Medical Data

Local on Premises/Cloud Database	Blockchain
Traditional record-keeping methods may not provide reliable auditing.	Blockchain's replicated, timestamped ledger facilitates efficient and trusted auditing.
Locally stored medical data might be unavailable in critical real-time situations.	Data replication ensures real-time availability from the local copy of the ledger.
Local on premises /cloud databases often lack proper authentication, leading to potential misuse of medical information.	Blockchain's encryption and digital signature techniques ensure user authenticity for accessing and uploading medical information.
Local on premises /cloud databases typically have less secure data access control.	Access control rights can be defined in smart contracts, ensuring secure and

Local on Premises/Cloud Database	Blockchain
	trustless data access for network participants.
In a Local on premises /cloud database, unauthorized users can impersonate legal users to access sensitive medical data.	The private blockchain network restricts data access to authorized participants based on access control rights.
Medical data in local databases can be easily altered or deleted.	Medical data are stored as transactions in blocks, linked cryptographically to ensure immutability.
Without robust security, users might deny accessing or modifying data in local on premises /cloud database systems.	Each operation's authenticity is recorded in an immutable ledger, preventing repudiation.
In a Local on premises /cloud database, unauthorized users can impersonate legal users to access sensitive medical data.	The private blockchain network restricts data access to authorized participants based on access control rights.

Results Analysis

In Table 6, we can observe the performance of Neo and Fabric in terms of throughput during the first test scenario. It is interesting to note that Neo initially experienced an increase in throughput, reaching a peak of 629 transactions per second (TPS) at the 60-second mark. However, it subsequently declines to 304 TPS and remains at that level for the remainder of the test. Similarly, Fabric follows a comparable pattern, starting with a high throughput of approximately 800 TPS and gradually decreasing to around 450 TPS by the end of the test. This indicates that both platforms exhibit fluctuations in their transaction processing speeds throughout the test.

Table 6. Throughput comparison for Hyperledger Fabric and Neo blockchain in the first test scenario

Transaction duration	30sec	60sec	90sec	120sec
Neo Blockchain	438	629	304	329
Hyperledger Fabric	798	445	409	435

Regarding latency, it is worth noting that Neo consistently takes approximately 14 to 24 seconds to confirm a transaction throughout the test duration.

Table 7 illustrates the latency of Fabric in the first test scenario. Here, we can observe that as more transactions were submitted, the latency gradually increased. This implies that as the workload on Fabric intensified with a higher number of transactions, the time taken to process and confirm each transaction also increased.

Table 7. Latency comparison for Hyperledger Fabric in the first test scenario

Transaction duration	30	60	90	120
Hyperledger Fabric	0.05	0.12	0.1	0.1

Moving on to the second test scenario (Table 8), where we increase the number of transactions sent over the network, Fabric outperforms Neo once again. Fabric demonstrates an upward trend in throughput as the number of transactions increases, while Neo exhibits more fluctuations in its performance.

Table 8. Throughput comparison for Hyperledger Fabric and Neo Blockchain in the second test scenario

Transaction send rate	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
Neo blockchain	373	427	489	422	534	543	434	424	409	474
Hyperledger	426	470	630	577	676	707	718	732	740	754

Fabric										

During the second test scenario (Table 9), the latency for Fabric remains relatively stable, with a slight increase observed. The latency values consistently range between 0.03 seconds and 0.04 seconds throughout the test. This indicates that Fabric can maintain low and consistent latency even as the number of transactions increases.

Table 9. Latency comparison for Hyperledger Fabric in the second test scenario

Transaction send rate	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
Hyperledger Fabric	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.04

Delays in fraud detection can slow the identification of fraudulent activities, causing financial losses and putting patients' health at risk. High latency impedes the prompt discovery of fraud, giving wrongdoers the chance to persist in their schemes unchecked, possibly resulting in more harm and greater financial damage. Thus, reducing latency is essential to improve the efficiency and precision of fraud detection, ultimately protecting healthcare resources, patients' health, and the credibility of healthcare services [CITATION Set21 \l 1033]. Throughput and latency can be significantly impacted by large transaction sizes, which are driven by extensive file requirements and block size. In private blockchain networks, the computational complexity and energy consumption of encryption and decryption operations add to this burden. Furthermore, replicating the ledger across all nodes increases computational and network overhead, resulting in high energy consumption, low transaction throughput, and limited scalability. As the number of nodes increases, so does the volume of data transferred, leading to longer processing times. Additionally, the choice of consensus mechanism affects scalability; for instance, Proof of Work (PoW) is particularly known for its high energy consumption, further exacerbating these challenges.

According to Table 10, the CPU and memory usage comparison between Fabric and Neo reveals that Fabric utilized fewer resources compared to Neo. Throughout all the conducted tests, both platforms achieve a 100% success rate, indicating their reliability for securely sharing sensitive healthcare insurance data. Fabric consistently outperforms Neo in all the tests, showcasing its superiority in creating healthcare insurance fraud detection systems. The lower resource consumption by Fabric suggests that it offers more efficient resource utilization, making it an optimal choice for healthcare insurance fraud detection applications.

Table 10. Memory and CPU consumption comparison for Hyperledger Fabric and Neo Blockchain

	Memory (MB)	CPU (%)
Hyperledger Fabric	191	47%
Neo Blockchain	515	39%

Limitations

Validating the system with real-world healthcare data presents several challenges and ethical considerations. Obtaining access to real-world healthcare data can be difficult due to stringent regulations and privacy concerns. Ensuring the data is accurate, complete, and representative of the broader population can be challenging, as inconsistent or incomplete data can affect the validity of the results. Integrating diverse data sources and formats into a cohesive system requires significant technical expertise and resources. From an ethical standpoint, protecting patient confidentiality is paramount, necessitating robust measures to ensure data is anonymized and secure. Additionally, there is a risk of introducing bias if the data is not representative of the entire population, potentially leading to skewed results and harmful recommendations.

In addition to the technical and ethical considerations, the successful implementation of this system

requires the cooperation and acceptance of various participants in the healthcare system[CITATION Lei20 \l 1033]. Key stakeholders such as hospitals, clinics, insurance companies, and healthcare providers must be willing to contribute their data and support the integration of a blockchain-based solution. Each participant has unique requirements for data security, privacy, and interoperability that must be addressed to ensure their cooperation. Furthermore, patient data must be governed and secured to ensure privacy and controlled access. Ensuring that only authorized personnel can access sensitive information and that all data handling procedures comply with relevant regulations[CITATION Hsi20 \l 1033].

Conclusion and Future Work

Healthcare insurance fraud detection is crucial for the healthcare industry. This is due to the high costs incurred from healthcare insurance fraud. In addition, some frauds are at risk to patient health. In this paper, we design and implement smart contracts to detect healthcare insurance fraud. This is based on our proposed taxonomy of fraud scenarios. Furthermore, we utilize a blockchain platform that is specifically suited for healthcare insurance fraud detection. To improve the selection of a suitable platform, we design and implement a decision-map-based recommender system, which automates and streamlines the platform selection process. To feed the recommender system with suitable candidates, we propose a taxonomy of 102 blockchain development platforms. Through these efforts, we aim to improve the efficiency and accuracy of healthcare insurance fraud detection by leveraging the capabilities of blockchain technology. The recommender system reveals Fabric and Neo as the top 2 platforms candidates for the development of healthcare insurance fraud detection, with Fabric having the highest rank. Our experimental numerical evaluation of the two selected platforms shows that Fabric outperformed Neo, demonstrating a more suitable network structure and features than Neo. Furthermore, based on our experiments, Fabric offers greater configurability, enabling further performance improvements. Machine and deep learning algorithms are alternative promising approaches for detecting patterns of fraud in large-scale environments such as healthcare insurance. However, these algorithms suffer from security and privacy issues and are prone to bias. In future work, we aim to explore machine learning techniques integrated in blockchain towards privacy-preserving machine learning models. In particular, we will explore how can blockchain improve the privacy and security of machine learning models, investigate the most effective ways to integrate federated learning with blockchain to ensure data remains decentralized and secure, and design smart contracts to automate and verify the training and deployment of machine learning models.

References

CITATION 235 \l 1033 : , (1),
CITATION USD23 \l 1033 : , (2),
CITATION Hea23 \l 1033 : , (3),
CITATION Ala22 \l 1033 : , (4),
CITATION Roh21 \l 1033 : , (5),
CITATION Ire21 \l 1033 : , (6),
CITATION KHY22 \l 1033 : , (7),
CITATION Din18 \l 1033 : , (8),
CITATION Lei201 \l 1033 : , (9),
CITATION Ism202 \l 1033 : , (10),
CITATION Ism191 \l 1033 : , (11),
CITATION Lei20 \l 1033 : , (12),
CITATION Ism203 \l 1033 : , (13),
CITATION Lei211 \l 1033 : , (14),
CITATION Far20 \l 1033 : , (15),

CITATION Jia23 \l 1033 : , (16),
CITATION Sow19 \l 1033 : , (17),
CITATION Set23 \l 1033 : , (18),
CITATION Set23 \l 1033 : , (18),
CITATION Sow19 \l 1033 : , (17),
CITATION Jia23 \l 1033 : , (16),
CITATION Mac20 \l 1033 : , (19),
CITATION Sal20 \l 1033 : , (20),
CITATION Wei19 \l 1033 : , (21),
CITATION Lei21 \l 1033 : , (22),
CITATION Ism192 \l 1033 : , (23),
CITATION Rim23 \l 1033 : , (24),
CITATION Sat08 \l 1033 : , (25),
CITATION Eth \l 1033 : , (26),
CITATION Ism191 \l 1033 : , (11),
CITATION Hyp \l 1033 : , (27),
CITATION NEO \l 1033 : , (28),
CITATION Xin21 \l 1033 : , (29),
CITATION Con \l 1033 : , (30),
CITATION Lei21 \l 1033 : , (22),
CITATION Mar22 \l 1033 : , (31),
CITATION Lei08 \l 1033 : , (32),
CITATION Raj23 \l 1033 : , (33),
CITATION Hyp4 \l 1033 : , (34),
CITATION Neo \l 1033 : , (35),
CITATION Lei211 \l 1033 : , (14),
CITATION Set21 \l 1033 : , (36),
CITATION Lei20 \l 1033 : , (12),
CITATION Hsi20 \l 1033 : , (37),

Abbreviations

ABFT	Asynchronous Byzantine Fault Tolerant
BFT	Byzantine Fault Tolerance
DAG	Directed Acyclic Graph
DBFT	Delegated Byzantine Fault Tolerance
DPoS	Delegated Proof of Stake
FPC	Fast Probabilistic Consensus

IBFT	Istanbul Byzantine Fault Tolerance
NR	Not reported
PBFT	Practical Byzantine Fault Tolerance
PoA	Proof of Authority
PoC	Proof-of-Capacity
PoET	Proof of Elapsed Time
PoH	Proof of History
PoO	Proof of Ownership
PoP	Proof-of-Pledge
PoR	Proof of Replication
PoR	Proof-of-Randomness
PoS	Proof of Stake
PoSA	Proof of Staked Authority
PoX	Proof of Transfer
PoX	Proof of Space
PPoS	Pure Proof of Stake
RBFT	Redundant Byzantine Fault Tolerance
SBFT	Smilo Byzantine Fault Tolerance
SCP	Stellar Consensus Protocol
YAC	Yet Another Consensus

Supplementary Files

Untitled.
URL: http://asset.jmir.pub/assets/d5cf6b70a40bffd2e6d74c6ee3d02285.docx

Untitled.
URL: http://asset.jmir.pub/assets/7e3abc8cb9f02afbd94ec592959aa89b.docx