IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Blockchain and AI-empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects

KHYATI KAPADIYA[1], USHA PATEL[1], RAJESH GUPTA[1], STUDENT MEMBER, IEEE, MOHAMMAD DAHMAN ALSHEHRI[2], SUDEEP TANWAR[1], SENIOR MEMBER, IEEE, GULSHAN SHARMA[3], PITSHOU N BOKORO[3]

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India (e-mails: 20mcec19@nirmauni.ac.in, ushapatel@nirmauni.ac.in, 18ftvphde31@nirmauni.ac.in, sudeep.tanwar@nirmauni.ac.in)
[2]Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia (e-mail: alshehri@tu.edu.sa)
[3]Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa. (e-mails: gulshans@uj.ac.za, pitshoub@uj.ac.za)

Corresponding author: Sudeep Tanwar (e-mail: sudeep.tanwar@nirmauni.ac.in) Usha Patel (ushapatel@nirmauni.ac.in),

Taif University Researchers Supporting Project number (TURSP-2020/126), Taif University, Taif, Saudi Arabia

ABSTRACT Nowadays, health insurance has become an essential part of people's lives as the number of health issues increases. Healthcare emergencies can be troublesome for people who can't afford huge expenses. Health insurance helps people cover healthcare services expenses in case of a medical emergency and provides financial backup against indebtedness risk. Health insurance and its several benefits can face many security, privacy, and fraud issues. For the past few years, fraud has been a sensitive issue in the health insurance domain as it incurs high losses for individuals, private firms, and governments. So, it is essential for national authorities and private firms to develop systems to detect fraudulent cases and payments. A high volume of health insurance data in electronic form is generated, which is highly sensitive and attracts malicious users. Motivated by these facts, we present a systematic survey for Artificial Intelligence (AI) and blockchain-enabled secure health insurance fraud detection in this paper. This paper presents a taxonomy of various security issues in health insurance. We proposed a blockchain and AI-based secure and intelligent system to detect health insurance fraud. Then a case study related to health insurance fraud is presented. Finally, the open issues and research challenges in implementing the blockchain and an AI-empowered health insurance fraud detection system is presented.

INDEX TERMS Healthcare insurance, Fraud detection, AI, Blockchain, Security

I. INTRODUCTION

Health insurance (HI) is a contract between the insurance provider and insurance subscriber in which the provider compensates the insurance subscriber's healthcare expenses. The Health insurance association of America stated that healthcare insurance covers losses resulting from accidents, healthcare expenses, incapacity, accidental injury, and damage [1]. Insurance subscribers have to pay the premium regularly for this compensation. The insurance provider can be from the commercial world or a government body. Nowadays, HI has become a necessity for each individual due to the rising hospitalization and treatment costs and getting income tax rebates [2]. Earlier, the health insurance claim (HIC) process was manual and offline, with many shortcomings, such as insurance subscribers need to visit the insurance office during office hours only to fill out the premium and inquire about the HIC status, which wastes time and money in terms of transportation costs. This procedure is wholly based on pen-paper, so human resource necessity and the possibility of error are more for auditing HIC. Maintaining and integrating the paper-based health claim data is very tedious and challenging work. Health claim records are easily alterable and accessible. So, the chances of fraud occur from the insurance provider, insurance

subscriber, and healthcare service provider due to lesser transparency and privacy. It is less cost-effective due to the involvement of the intermediary broker or agent costs [3].

In the digital era, every piece of information is gathered in a digital form, which revolutionizes the HIC worldwide. Following are various benefits of digitization: (i) it provides convenience to the parties involved with HIC. Insurance subscriber does not need to visit the insurance office frequently to purchase an HI and to fill out the premium amount [4], (ii) communication between insurance subscribers and providers becomes efficient, (iii) it makes auditor's complex and tedious work easy, (iv) any kind of fraudulent behavior can be easily identified using AI, (v) it also reduces the human resource cost [4] and (vi) verification of claims becomes fast using web-generated reports, so insurance subscribers get insurance coverage fast and automatically during any medical emergency. There are many other benefits of the digitization of HIC besides those mentioned above.

Despite several benefits of digitized HIC, it faces various challenges, which are described as follows.

- *Validation of data and model*: Insurance providers use a digital business model to determine the premium rate and coverage price. This model is developed by professionals who may be unfamiliar with the rules and specific requirements of HIC, which can make professionals find it challenging to measure the impact of new variables used in the models [5]. In web-based HIC, data is collected from various social media platforms that may be no 100% accurate and outdated for validating the digital business model [5]. The subscriber does not provide the data generated from the social media platform to the insurance provider. The insurance subscriber does not have a chance to correct this social media platform data, which may be used to determine the premium rate.
- *Lack of talent*: Developing web-based HIC needs a group of skilled people because it is dependent on complex algorithms and mathematical skills [6]. Lack of talent leads to the insurance sector becoming expensive.
- *Fraud detection system*: Every insurance plan, including HI, is vulnerable to fraud. Every year, HI provider firms lose revenue due to fraudulent claims. Insurance firms hike premiums to maintain profit, which impacts legitimate insurers. It is assessed that fraudsters approximately take fifteen percent of the taxpayer's money, which is utilized to finance government-assisted medicare. So, this is necessary for national authorities to develop systems for detecting fraudulent cases and payments. HI fraud is a severe offence that affects people and the nation's money and time. As a result, a good fraud detection system is necessary for lowering costs and enhancing the security of healthcare [1]. HI fraud is increasing daily, which is a concern for the nation's insurance subscribers and providers. As there is no option for a HI fraud penalty, the number of fraudulent cases is increasing as a rapid rate [6].
- *Connecting to outdated computer system*: Insurance providers need to replace legacy computer systems or customize their systems to interface with new technologies. However, such technologies increases the back end cost for he insurance provider, but provides security against the HIC fraud. Legacy computer systems was developed around satisfying regulatory requirements rather than enhancing the subscriber's experience [5].
- *Privacy of HI subscriber's data*: The insurance providers expanded the use of subscriber's data, which raises concerns about its security and privacy. It can not be possible for an insurance subscriber to know exactly what or when data is collected and how data is being used. So, insurance subscriber does not provide explicit consent to the insurance provider, and insurance subscriber loses control over their personal information [5]. So, the identity threat is a big issue in which insiders can misuse the insurance subscriber's identities to get insurance coverage.
- *Security of HI system*: Digitization in HI raises various security concerns such as ransomware attacks, phishing attacks, Distributed Denial of Service (DDoS) attacks, replay attacks, and many more. Cybercrime affects the HIC industry from both internal and external sources, including the third parties [7]. HIC data is stored in various systems, and it is interlinked between systems which causes authentication and authorization problems. Insurance firms lose lots of revenue and reputation due to the compromised security of the HI system.

To overcome the aforementioned security issues, researchers have given various cryptography-based solutions. For example, Lou *et al.* [8] presented an access control mechanism for patient data based on attribute encryption. This data is shared among insurance firms, hospitals, and patients. The encryption technique is one-to-many, and ciphertext can be accessed by a group of people who fulfil certain access policies. Then, Heurix *et al.* [9] discussed pseudonymization and personal metadata encryption techniques for handling and exchange of EHR, personal health records, and billing records of patients. Usage of the encryption technique prevents unauthorized data disclosure and assures data security from internal attackers by decreasing sensitive data leakage. Later, Kumar *et al.* [10] developed an encoding algorithm for storing patient's private data on a server. Admin can access the patient and hospital-related information. Encoding of data is done by implementing the 128 base encoding method, and decoding is done with the help of the base64 decoding method. Later, Bellisario *et al.* [11] utilized an X.509 certificate for secure communication between insurance and healthcare providers and the authentication server. According to the hash value, they encrypted the EHR with a 256-bit AES key. The first-

time hash value is stored in the encrypted configuration file for future use, giving centralized access to EHR.

The aforementioned solutions offer security to the HIC data, but it still has several issues. The central server could crash due to a malicious attack or fault, which affects HIC business operation and work location availability [11]. Another disadvantage is immutability, which means data can't be removed or altered after being stored. The traditional security solutions proposed by the authors do not provide immutability and transparency, leading to data tampering in the system [11]. Authentication of user's information depends on the admin, and if the admin's security is compromised, the entire system would be vulnerable to security attack [10]. Data redundancy is also a demerit because multiple copies of the same data of HIC exist in the database. The aforementioned problems of the security-based solution can be resolved with the help of blockchain.

Blockchain distributed ledger allows the group of peers to collaborate to form a decentralized network. Consensus mechanisms are used for communication and sharing data among peers [12]. Blockchains store all transactions on a public ledger in the form of chained blocks without a central server [13]. It is a decentralized network that enables access to HIC records to all the parties involved in the HI without the involvement of any central authority to oversee the global HI data. The data saved on the blockchain cannot be corrupted, altered, or retrieved to ensure immutability and transparency in the system [13]. It enhances data security by utilizing encryption techniques using cryptographic hash keys and timestamp protocols. The availability of HIC data saved on the blockchain is ensured since the records on a blockchain are duplicated in numerous nodes, and the system can be protected against security attacks on the HIC confidential data availability. It can verify the authenticity of blockchain data without accessing the plaintext of those records [13]. From the discussion of various issues of digitized HI, in this paper, we are primarily focused on HI's fraud, security, and privacy issues.

A. SCOPE OF SURVEY

Here, we discuss the existing literature and present a comparative-analysis of various existing surveys and the proposed survey on HIC fraud detection.

Many researchers have presented surveys on HI fraud detection. Most of these surveys have given insights into the processing and integration of HIC datasets, data mining, and machine learning (ML) techniques for fraud detection in HI. For example, Duman *et al.* [14] presented the big data analytics aspect in HIC fraud detection. Their study provides detailed information about the HIC data source, benefits, and characteristics of the algorithm and model used for HI fraud detection. Then, Thomas *et al.* [2] addressed the supervised learning technique for fraud detection systems that have been optimized in the HI industry to identify the fraudulent claims. Later, Bauder *et al.* [15] analyzed the various CMS medicare and List of Excluded Individuals and Entities (LEIE) datasets, which are being processed for identity fraud in HIC. CMS medicare dataset summarizes the information about healthcare services providers. It also gives information about medical procedures prescribed to HI beneficiaries and prescription drugs managed by the physicians under the Medicare part D prescription drug program. LEIE dataset contains information about the fraud labels. They also analyzed mapping between the LEIE and CMS datasets to identify the HIC fraud detection. Further, Ekin *et al.* [16] discussed statistical HIC fraud assessment using sampling, overpayment estimation, data mining technique, and the Bayesian approach. They also demonstrate various unsupervised learning techniques for fraud detection in HIC utilizing a real-time dataset.

Later, Ankrah *et al.* [17] presented an exhaustive survey on Ghanaian HIC dataset. They introduce the dataset feature in the context of ghana's healthcare system. This dataset was categorized into six groups, i.e., the information of the client, services offered, diagnosis, investigations, medicines, and client summary. Then, Chen *et al.* [18] described several types of HI fraud and explained recommendations and techniques for combating HI fraud. They also explored healthcare fraud cases of pharmacogenetic testing. Later, Mary *et al.* [19] investigated the imbalance classification issue that occurred in the method solution for the HI Fraud detection. In an imbalanced classification problem, data distribution across known classes is biased. The comprehensive research study shows that the class imbalance in the Support Vector Machine (SVM) classifier is more influenced than decision trees and neural networks classifiers. Further, Magalingam *et al.* [20] presented various data mining methods for fraud detection in finance applications such as insurance fraud, bank fraud, corporate fraud, and cryptocurrency fraud. They found SVM extensively utilized in fraud detection of financial applications from their survey. The research study also shows the list of nations vulnerable to financial fraud.

In this proposed review, we have discussed major security issues and their countermeasures in HIC and proposed a blockchain and AI-empowered architecture for HIC fraud detection. Table 1 shows a comparative analysis between the existing review and the proposed review on HIC fraud detection.

B. MOTIVATION

The motivation of this paper is as follows.

- The importance of security, privacy, and fraud detection in HI is key criteria. Without the security and privacy of the HIC system, patient's sensitive Personally Identifiable Information (PII) can be compromised, which can ruin the insurance firm's reputation. Fraud in healthcare insurance causes loss for individuals, private firms, and governments. So, the devise of secure fraud detection methods for HIC has become necessary.

TABLE 1: Comparative analysis between existing review and proposed review on HIC fraud detection.

Author	Year	Contribution	Pros	Cons
Duman et al. [14]	2017	A review on fraud detection techniques and big data analytics aspects in HI.	Provide significant HIC dataset related information	Only ML based solution discussed
Thomas et al. [2]	2018	Present a review on supervised learning technique's impact on HI fraud	Useful for labeled data	Not given information about tools
Bauder et al. [15]	2018	A survey on Medicare dataset integration and processing of fraud detection system	It guides researcher about research gap in medicare dataset processing and integration	Does not analyze Security of dataset
Ekin et al. [16]	2018	A review on statistical healthcare fraud assessment	It provides useful information for the complicated nature of healthcare insurance data and the heterogeneity of healthcare systems.	Not given detailed security aspect in statistical assessment
Ankrah et al. [17]	2018	A Review on HI claim dataset of Ghana	Helpful in drug utilization research aspect	Limited discussion about HI fraud detection from dataset
Chen et al. [18]	2020	A review on recommendations method for healthcare service providers and patients to combat Fraud	Provide awareness information about healthcare fraud	Lack of detailed information about fraud prevention method
Mary et al. [19]	2021	Study present class imbalance issue of ML method in detection of healthcare fraud	The researcher can easily identify affecting factor of class imbalance	Not discussed detail solution of class imbalance issue
Magalingam et al. [20]	2021	A Review on data mining technique for fraud detection in all types of finance application	Cover useful information about the dataset and validation measures for estimating the performance of data mining techniques.	Limited information about HI fraud
Proposed Survey	-	A survey on blockchain and AI-empowered HI Fraud Detection	Present security issues, solutions, and architecture for HI fraud detection	-

- In existing surveys, security issues and HI fraud detection were not discussed. So, there is a need for a comprehensive survey that inspects the secure AI and blockchain empowered HI fraud detection system.

C. CONTRIBUTION

This paper presents a detailed survey of HI fraud detection. We highlight various open issues and research challenges in HI fraud detection. Following are the research contributions of the paper.

- We present a background and various security and privacy issues of HI fraud detection and present a taxonomy on possible security attacks on the HI systems along with their countermeasure tools.
- We propose a blockchain and AI-based system to fight against various security issues in HIC fraud detection that increases the transparency and trust among the HI provider and subscriber.
- We also present a case study on HIC fraud detection using healthcare wearable devices.

D. ORGANIZATION

Figure 1 shows the organization of the proposed survey, which is as follows. Section 1 provides an introduction to HI. Section 2 describes the background knowledge of HI, HIC fraud types, and technology used to detect HIC fraud. Then, Section 3 discusses several security issues in HI. Section 4 presents the category-wise taxonomy of HI fraud detection. Later, we propose an AI and blockchain-based secure architecture of HI fraud detection in Section 5. Next, Section 6 present a case study using the proposed architecture. Section 7 confer the proposed work's open research issues and future directions and conclude the paper in Section 8. Table 2 shows various abbreviations used in the paper.

II. BACKGROUND

This section provides the background of HI and HIC fraud types and detection of HIC fraud.

A. REVOLUTION OF HI

HI is the most effective method of financing healthcare to preserve a person's money and prevent indebtedness. Healthcare insurance helps subscribers to pay for expensive healthcare services in case of medical emergency [19]. In several countries, healthcare is becoming a considerable expenditure. It is considered crucial for the life of several residents in a particular country. Many patients receive inadequate healthcare services due to unaffordable high medical service costs [21]. HI has already become an essential part of people's life. Healthcare spending has been steadily rising, so this has already become a worldwide phenomenon [22].

In India, HI has become an essential part of the economy. The availability of medical services in India differs by state. Most states have public medical services; however, due to a shortage of resources and administration, most citizens prefer private medical services [23]. India consists of a global

TABLE 2: Abbreviations

Abbreviations	Explanation	Abbreviations	Explanation
HI	Health Insurance	SMOTE	Synthetic Minority Over-Sampling Technique
HIC	Health Insurance Claim	LightGBM	Light Gradient Boosting Machine
PHI	Personal Health Information	PCA	Principal Component Analysis
LEIE	List of Excluded Individuals and Entities	DNN	Deep Neural Network
SVM	Support Vector Machine	BILSTM	Bidirectional Long Short-Term Memory
PII	Personally Identifiable Information	PNN	Parallel Neural Network
HIPPA	Health Insurance Portability and Accountability Act	SNN	Simulated Neural Network
EHR	Electronic Health Record	ECM	Evolving Clustering Method
DDOS	Distributed Denial of Service	BERT	Bidirectional Encoder Representations from Transformers
MITM	Man In The Middle	L-SVM	Lagrangian Support Vector Machine
API	Application Programming Interface	GRPC	Google Remote Procedure Call
ML	Machine Learning	NS 3	Network Simulator
DBN	Deep Belief Network	ICD	International Classification of Diseases
RIPPER	Repeated Incremental Pruning to Produce Error Reduction	DPCNN	Deep Pyramid Convolution Neural Network
HAN	Hierarchical Attention Network	TEXT-RNN	Text Recurrent Neural Network
IPFS	InterPlanetary File System	LEAM	Label-Embedding Attentive Model
POS	Proof of Stake	POW	Proof of Work

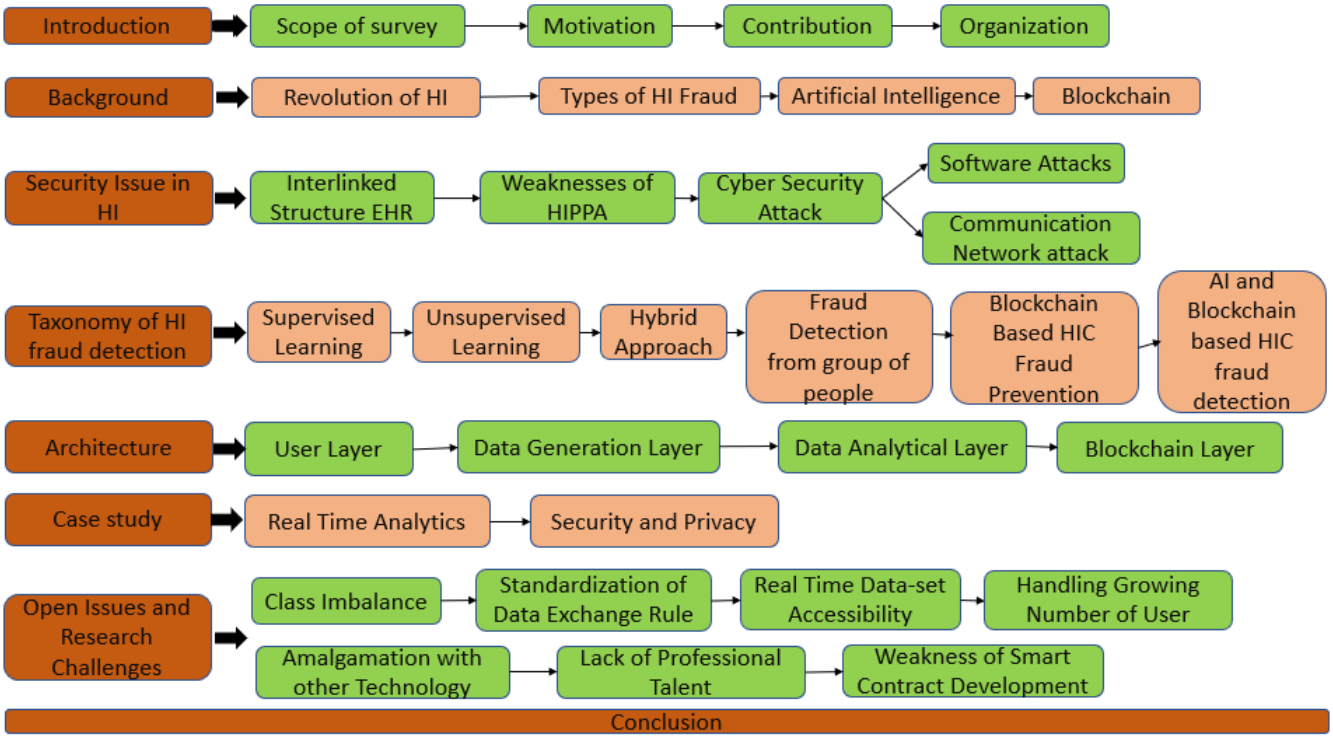

FIGURE 1: Organization of the survey.

and single-payer medical system financed by either public or private HI funds and a section of government medical centres that is almost entirely paid by taxes [23]. India has introduced a lot of revolutionary HI policies. FIGURE 2 shows the year-wise revolution in HI policies.

The Constituent Assembly approved the National Health Policy in 1983, which was further modified in 2002 and 2017 [23]. The update in policy is required due to the growth in diseases and rising incidences of unaffordable expenditure because of medical services costs. India's first medical plan was introduced in 1986 to regulate the terms and conditions of the HI [1]. It covered inpatient costs but excluded existing diseases/conditions such as pregnancy, maternity, and HIV. The payments were reimbursed automatically by third-party administrators. India's constitution introduces a new economic policy in which the insurance sector was privatized in 1991 [1].

The Indian government liberalized insurance in 2000, which allowed private businesses to enter the market. With the arrival of private health insurers in India, various novel policies and plans such as family floater plans, top-up plans, critical illness plans, and hospital cash were introduced [1].

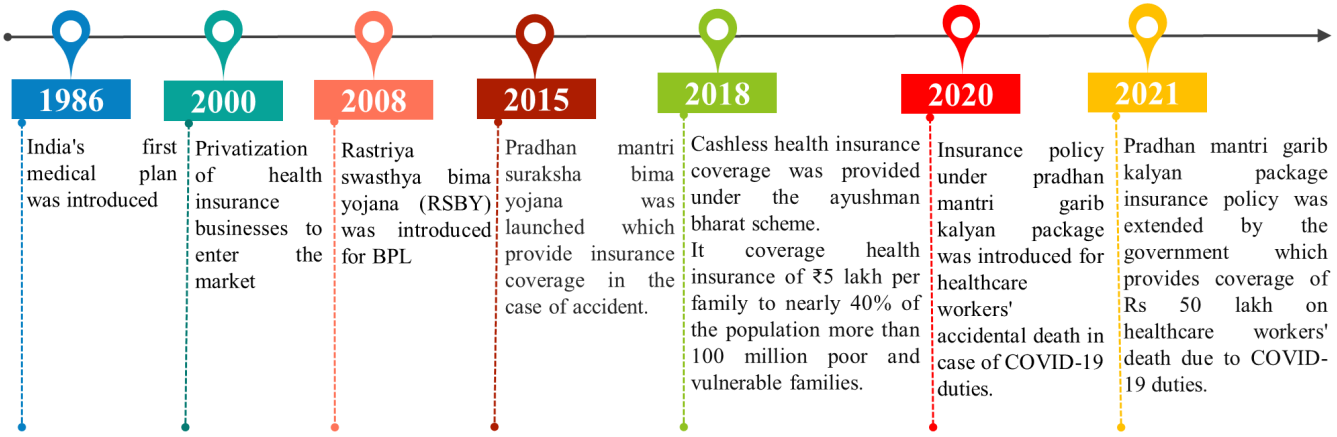

FIGURE 2: Revolutions in HI policies in Indian context.

In the year 2008, the Rashtriya Swasthya Bima Yojana was introduced, in which the HI plan aims to ease the progressive rollout of HI projects for BPL workers in the states of India [23]. Pradhan Mantri Suraksha Bima Yojana (PMSBY) was introduced in 2015. It is a government accident insurance policy in India. Insurance subscribers get coverage of Rs 2 lakh in unfortunate death and Rs 1 lakh in the incident of severe lifelong disability under this policy [24]. In 2018, the Indian government started Ayushman Bharat, a countrywide taxpayer-funded HI scheme. This program intends to address the lower half of the nation's population and provides individuals with free medical care in both public and private medical centres [1]. In 2020, an insurance policy was launched under Pradhan Mantri Garib Kalyan Package for healthcare workers fighting COVID-19. This policy provides the coverage of Rs 50 lakh to private and government healthcare workers who may have accidental death in case of COVID-19 duties [25]. This policy was launched for around 22.12 lakh healthcare providers. In 2021, the government extended the Pradhan Mantri Garib Kalyan Package Insurance policy, which provides insurance coverage to healthcare workers suffering from severe disease or death due to the COVID-19 duties.

B. TYPES OF HI FRAUD

Medical insurance fraud is a serious subject in each country and the forged behavior patterns vary according to the situation. Various types of HI frauds occur in each country and multiple parties are involved with this fraud. There are mainly three parties engaged in this fraud. First, healthcare service providers such as doctors, hospitals, ambulance firms, and laboratories. The second is subscribers such as patients and their employers. The third is HI provider's fraud which includes private insurance firm, and the government sector [14]. Table 3 shows the different types of HI fraud. For decades, fraud has been a significant problem and may be found in any industry. Every time, fraudsters used the new technique to commit fraud. Individuals, businesses, organizations, and governments may face severe losses due to healthcare fraud. It is also predicted that the $600 to $850 billion per year is lost because of frauds in the medicare system of United States and face loss of approximately $125 to $175 billion of the total amount due to the forged activities involved in the system [22]. So, combating fraud is becoming a priority for all nations. Several experts propose fraud detection methods in HI. Fraud can be attempted by fraud identification method at their occurrence. Manual and automatic fraud monitoring are the two most common procedures for detecting fraud. Monitoring fraud procedures requires exhausting human labour. This procedure contains complicated transactions which take more time and involve a wide range of field experience and expertise. As a result, automated techniques have been developed to enhance the effectiveness of fraud detection. In automatic fraud monitoring, various data mining computer-based methods are involved [14].

C. ARTIFICIAL INTELLIGENCE (AI)

Artificial Intelligence (AI) techniques have been utilized as a valuable tool for HIC fraud detection. AI automates the HIC fraud detection system. As per the recent studies, AI has been mainly used to solve HIC fraud detection using several ML, deep learning, and data mining models [29]. Behavioral profiling methods based on ML techniques are used to detect anomalies and fraud detection. For this purpose, each individual's behavior pattern is modeled to monitor it for any derivation from norms [14]. ML techniques used in HIC fraud detection are categorized into supervised learning, unsupervised learning, and semi-supervised learning [14].

Supervised learning technique in HIC fraud detection uses a dataset of previously known fraudulent and legitimate records. Those records are utilized to capture fraud patterns and build the model. The essential benefit of the supervised learning technique is that the classification results given by this technique are easy to comprehend. Various classification and supervised learning regression analysis algorithms are used in HIC fraud detection. Neural networks, SVM,

TABLE 3: Types of HI frauds.

Healthcare service provider's fraud	HI subscriber's fraud	HI provider's fraud
Medical services that have not been performed are being billed [26].	To get a lesser premium rate, misrepresent eligibility documents [14].	Charge of premium taking more from insurance subscriber by falsifying subscriber's claim [27].
Expensive services and medical tests are being billed that are costlier than the original test [26]	Making claims for medical treatments that were never provided [14]	Claims are being blocked without an examination of the claims' authenticity [28].
For the goal of gaining insurance coverage, presenting non-covered therapies as medically not required [26].	Obtaining financial compensation by impersonating another person's policy card [26].	To discourage the insurance subscriber, the HI provider incorrectly dismisses legitimate claims in the expectation that the patient will soon quit [28].
Misrepresentation of a specific diagnosis or history of therapy [14].	Actively participating with scam networks through buying several insurance policies [26].	Useless and forged claims are created by insurance providers for collecting premiums from insurance subscribers [28].

decision trees, bayesian approaches, statistical analysis, graph analysis, and rule-based methods are supervised learning techniques used to identify HIC frauds [30] [31]. Apart from benefits, supervised learning has many drawbacks. In supervised learning techniques, data collection and data generation are challenging. If the dataset of HIC is quite huge, then the labelling of data is a bit difficult. When labels are uncertain and ambiguous, it is difficult to identify between them in fraud data [14]. In some circumstances, supervised learning implementation becomes challenging due to these constraints. These constraints of supervised learning can be solved with the help of unsupervised learning.

Unsupervised learning techniques identify the fraudulent behaviors of HIC from the unlabeled HIC datasets. The benefit of HIC fraud detection using unsupervised learning is that there is no need for labelled data [14]. Unsupervised learning is used in HIC fraud detection where labelled data is unavailable. Various unsupervised learning techniques are utilized in HIC fraud detection, such as association rules mining, data mining, k-means clustering, and k-nearest neighbor. Semi-supervised learning is used to get benefits from both supervised and unsupervised learning. These hybrid semi-supervised learning are used where a limited number of labelled data is available compared to unlabeled data. A predictive model is built using both labelled and unlabeled data in semi-supervised learning [14]. Combined clustering and classification as a semi-supervised learning method is generally used in HIC fraud detection.

D. BLOCKCHAIN

Security is one of the key concerns in HI. Existing centralized systems provide security to a certain extent, but they could crash due to malicious attacks or faults. This issue can be solved with the help of a decentralized network, i.e., blockchain. The idea of blockchain was discovered in Satoshi Nakamoto's proposal for the virtual currency, i.e., bitcoin. A blockchain is a digital ledger of transactions stored in a chain of blocks [32]. A cryptographic hash of the previous block, a timestamp, block header, block number, version, nonce, and transaction data are all included in each block. Transaction data is maintained as a Merkle tree [33]. In blockchain, if any of the transactions in a block are modified or altered, leading to the drastic change in the hash value of the particular block. It further breaks the chain of blocks on the blockchain, which helps to detect the modified transaction. Hence, a transaction can't be modified or altered once added to the blockchain. As a result, data on the blockchain is immutable. Every transaction has been digitally signed with a private key utilizing asymmetric cryptography to assure integrity, security, and immutability [33]. Only the public key owner can authenticate the signing entity. These signatures prevent tampering with the transaction data [34].

The blockchain communication network is decentralized and peer-to-peer, which solves the problem of a single point of failure. Blockchain utilizes a consensus protocol instead of a central authority to settle disagreements between nodes in a distributed application [33]. A consensus protocol is used in the blockchain networks to ensure reliability and trust between unknown peers in a distributed computing environment. In the consensus protocol, all peers in the blockchain network agree on the distributed ledger's current state. There are four types of blockchain: public, private, consortium, and hybrid [35] [36]. In a public blockchain, everyone with internet connectivity can sign on the blockchain to become an authorized node making the public blockchain permissionless [37]. Permissions are required to use a private blockchain. It is only open to those who have the network administrator's permission. A hybrid blockchain combines the benefits of both centralized and decentralized blockchain [37]. A consortium blockchain contains the combined features of a private and public blockchain. There are two types of nodes in the blockchain based on their role, i.e., whether it works as a full node or as a normal node. A full node(mining node) keeps a copy of the entire transaction record and participates in the validation and authentication process, including signature verification and mining [33]. In a blockchain network, full nodes are the backbone of trust. Because it is in charge of implementing consensus norms and processes, the normal node could create and transfer transactions while maintaining the blockchain's header. It can not participate in the validation process.

Blockchain as a decentralized framework is beneficial to HIC applications which allow operations of distributed HIC applications without relying on a centralized server. Replication of blockchain ledger data across all the nodes generates an environment of transparency among all the parties involved in the HIC process [37]. The attack that happens on any one node in the blockchain network does not affect the ledger's status due to the data in the blockchain ledger being replicated across all nodes in the network. One of the HIC fraud detection system requirements is the storage of HIC data that are not alterable [37]. The immutability feature of blockchain fulfils aforementioned requirement. Preserving the integrity and authenticity of the insurance subscriber's (patient) records is crucial. Insurance subscriber's data on the blockchain is encrypted using cryptographic techniques, which assure that only insurance subscribers with genuine authorization can access and decrypt the data. It further enhances the data security and privacy of the HIC. Insurance subscriber's data can be exchanged among all parties involved with HIC fraud detection without revealing the insurance subscriber's identities because the identities of insurance subscriber's in a blockchain are pseudonymized using cryptographic keys [13]. Blockchain supports smart contract, which is written in the form of an executable code, and designed in solidity language, and deployed on Ethereum Virtual Machine to regulate activities following the agreement [13]. It contains transaction logic that governs the whole HI business process. Smart contracts are executed once the transaction commits the operation. A smart contract defines rules for the HIC process that allows insurance subscribers to control how their HIC records are shared or used in the network [13].

III. SECURITY ISSUES IN HI

The high volume of healthcare data in electronic form is generated due to technological advancements. This data is very sensitive and susceptible to security attacks. So, security is a major issue in HIC. The following are major security issues in the HIC.

A. INTERLINKED STRUCTURE OF ELECTRONIC HEALTH RECORD (EHR)

The EHR contains sensitive data related to patient medical information. Physicians, healthcare workers, and insurance firms share crucial healthcare data of patients using EHRs. This will be easier to manage medical services and deal with insurance issues, but this interlinked structure of EHRs creates a security problem. Because of the interlinked structure of EHRs, attackers can access this record which has been gathered for years under patient's identities. Patient records must be shared to provide better treatment, yet this will make networks vulnerable [38].

B. WEAKNESSES OF HI PORTABILITY AND ACCOUNTABILITY ACT (HIPPA)

HIPPA was established for the privacy and security of healthcare data in the USA in 1996. HIPAA's rule covers how to use and release healthcare data and personal information. This rule ensures the security of personal data while granting the flow of healthcare data, which is needed for medical choices. HIPPA rule applies to the HIC and healthcare service providers who share healthcare data with HI provider [39]. HIPAA omits Fitbits, Apple watches, fitness, and fertility applications. This rule does not protect the enormous amount of data that people generate through their usage of digital devices, apps, e-commerce sites, and social media. HI firms, healthcare service providers, and insurance subscribers can easyily access to this data [40].

C. CYBER SECURITY ATTACKS

The Healthcare system provides life-saving care to patients using advanced technology. A high volume of healthcare data in electronic form is generated. This data is highly sensitive, and attracts hackers for cyber attacks. This healthcare data is used in HIC. The network of HIC is also vulnerable. It is an open channel for communication, so there is a possibility of network attacks on the data. Various attacks can be possible on HIC to breach security. We categorize possible security attacks on HIC into 1) software and 2) communication network attacks. The software attacks are a) Phishing b) Ransomware [41] c) SQL Injection d) Malware and e) Bruteforce. Software attacks are performed on the back-end of the HIC software to change its control and operations. The communication network attack are: a) Man in the Middle b) DDoS c) Evesdroping d) Replay and e) Impersonation. Communication network attacks are performed on the HIC network to obtain unauthorized access of HIC data. Table 4 describes possible security attacks and consequences on the HIC category-wise.

1) Software Attacks

In this type of attack, attackers design malicious programs to harm HIC system and servers. Utilization of software has completely transformed in the healthcare insurance industry, which undoubtedly helps insurance subscribers and providers. However, there is no adequate security measure to ensure that HI software functions in the authentic manner. As a result, HIC systems are vulnerable to various software and app-related risks, including phishing, Ransomware, SQL injection, malware, and brute force. The countermeasure tools like netcraft, open SSL, iperf tool, SQL rand, droidMat, specter, iptables, etc are used to protect the confidentiality, integrity, and availability of HIC data and prevent from software attacks. Machine learning and deep learning technology are used to prevent software attacks. Machine learning algorithms such as naive bayes, sequential covering algorithm, random forest, decision tree, clustering, etc are utilized for the countermeasure of software attacks. FIGURE

3 shows possible software attacks on HI and the associated countermeasure tools, technologies, and algorithms.

2) Communication Network Attacks

Patient's remote monitoring, diagnosis, treatment, and emergency support are all possible with rapid development in wireless communication technologies. A patient's medical data is generated from wireless communication, which various users share and access, including healthcare professionals, researchers, government organizations, and insurance providers. Attacks on wireless communication seem to be a key concern for the healthcare system. Attackers use eavesdropping, replay, impersonation, and DDoS to compromise the system's integrity and authenticity. Tools for the countermeasure of communication network attacks are burp, ettercap, trinity, tcpflow, tcpick, snort, tcprewrite, netcut, etc. Machine learning algorithms such as random forest, DNN, SVM, CNN, decision tree, reinforcement learning, etc are utilized to countermeasure of communication network attacks. Various cryptographic countermeasure algorithms such as symmetric-key cryptography, homomorphic encryption, cryptography hash algorithm, etc are used to prevent from communication network attacks. FIGURE 4 shows the possible communication network attacks on the HI and the countermeasure tools, technologies, and algorithms.

Communication network attacks are classified into active and passive attacks according to the attack's involvement in the HIC system. In an active attack, the attacker attempts to modify the content of the communication message. DDoS, the man in the middle (MITM), and replay attacks are active attacks. Active attacks on the HIC system compromise its integrity and availability of the HIC system. In a passive attack, attackers may eavesdrop to analyze and replicate the communication and use it maliciously. Eavesdropping and impersonation attacks are passive attacks. The passive attack can compromise the confidentiality of the HIC system.

We address the several countermeasure tools, technologies, and algorithms offered by the researchers to protect against the possible attacks described in the table. For example, Kok *et al.* [42] proposed a two-stage pre-encryption ML algorithm to detect ransomware. In the first phase, they examined the malicious program's application programming interface (API) using ML. This method was employed to guarantee the detection of both existing and undiscovered crypto-ransomware. In phase two, the signature repository is created in which the signature of the predicted crypto-ransomware is generated. The signature repository efficiently identifies crypto-ransomware using the signature matching method at the pre-encryption stage. Then, Singh *et al.* [43] propose a SQL injection detection technique using a clustering ML algorithm and identify the unauthorized users by maintaining an audit record. An audit record is being used to compare the attack intensity and detection probability to estimate the detection accuracy. Later, Aleroud *et al.* [44] classify the phishing countermeasures based on ML, text mining, human users, profile matching, ontology, honeypot, search engine, and client server-based authentication. Then, Datta *et al.* [45] suggested the malware countermeasure for the android platform. They suggested an android security extension that gives extremely quiet security controls over the android apps. It enhances the android authorization model by giving control over which type of contacts access and which sites an app can connect to over the internet. Later, Sadasivam *et al.* [46] proposed honeynet architecture for the detection of distributed brute force attacks. They showed the attacker's login attempts determined by the behavioral index. All the single-source attacks are identified according to the behavioral index.

Later, Shyam *et al.* [47] proposed a SaaS framework for DDoS attack mitigation based on a Deep Belief Network (DBN). The weight and activation function of the DBN is fine-tuned using the median fitness sea lion optimization technique. When DBN identifies an attack node, control is passed to a lightweight bait technique that successfully reduces the most frequent attack. Then, Zagrouba *et al.* [48] explained data integrity, routing security, and data privacy countermeasure for MITM attack. For data integrity and routing, security authentication and encryption method are useful. Cryptographic hash functions are the greatest defense from data integrity attacks.

Later, Yousefpoor *et al.* [49] reviewed several eavesdropping attack's countermeasure encryption methods which ensure the data privacy and data confidentiality across a wireless sensor network. They discussed the homomorphic encryption method, enabling the sensor nodes to secure their confidential information. An attacker will not decrypt data packets if he obtains the private keys of all sensor nodes. As a result, if an attacker eavesdrops on a communication channel, it can only see the encrypted data packets without accessing the complete information. Then, Waqas *et al.* [50] presented a reinforcement learning-based method to secure the system against impersonation attacks. They used channel gain to identify impersonation and generated valid secret keys between authorized users. Later, Bruce *et al.* [51] proposed a security solution for e-healthcare that mitigate replay and impersonation attacks. They also designed a protocol to ensure the system's security against impersonation attacks. If an adversary tries to steal the certificate to access the network, then the designed protocol can be used to verify the authenticity and take action accordingly. After the authentication procedure, the interacting entities generate a session key. This key is unique to each session and cannot be used again once it has ended, preventing replay attacks.

IV. TAXONOMY OF HI FRAUD DETECTION

In this section, we categorize the HIC fraud detection techniques. In the literature, various possible solutions for fraud detection in HI was given, which is discussed in the following section.

TABLE 4: Possible security attack On HIC

Security attack	Description	Consequence
Phishing	• Attackers send mass amounts of email with malicious links to users or employees of HIC for getting credential information of the user [44]. • It was built to look like emails coming from trusted and known parties.	• Disclose user's PII information and HI data,After getting credential detail, launch malware
Ransomware	• Disrupts a computer's sensitive functionality by encrypting a user's computer's sensitive file until he gets ransom money [52].	• Disrupt insurance business operations and harm the reputation of the firm
SQL Injection	• Attacker uses SQL to execute malicious code into a server, the server is forced to disclose secured data [53]. • HIC website comment section and search bar are used to execute malicious code .	• Attacker gets unauthorized access to the HIC database through SQL injection [53]. • Database contains sensitive PII information. The attacker can perform malicious activity and fraud in HIC with this information.
Malware	• Malware is a malicious software program written to interrupt a computer, server, or communication network [54]. • Malware-infected HIC systems might diverge from their intended functions, such as delaying or shutting down.	• An attacker can disclose private information of the HIC system and obtain unauthorized access to information or systems, deny users access to information, or inadvertently compromise a user's computer security and privacy.
Bruteforce	• Attackers get credentials of the HIC system by guessing all combinations and getting unauthorized access to the system [55].	• Attackers steal PII-related data and execute malware to the network through unauthorized access. HI firm's reputation is also ruining
Man-in-the-Middle	• Attackers intercept communication between server and client, and the victim is clueless about the attack. • This attack can happen between parties involved in HIC such as HI subscribers and insurance providers [56].	• An attacker obtains confidential information from the victim by hiding identity.
DDOS	• Attackers overload the server with fraudulent service requests containing fraudulent return addresses, distracting the server during authenticating service requests [57]. • Multiple attacker machines targets one machine for attack.	• It causes server down and does not enable service request processing. • System goes completely offline. This attack harms the insurance firm's service. Also, the firm has to suffer an expanse of recovery.
Eavesdropping	• Attacker attempts to steal HIC data from a smartphone while the user is sending or receiving data over a communications channel by exploiting the vulnerability of the communication channels [49]. • On a computer or a server, the attacker deploys a network monitoring tool sniffer to capture data as it is transferred.	• The attackers can be sold important corporate and financial data of HIC for malicious activities.
Replay	• A replay attack occurs when an attacker intercepts the HIC network connections and falsely resends them to misdirect the receiver.	• If an attacker eavesdrops on HI firm's financial transaction through a replay attack, the firm will lose money.
Impersonation	• The attacker masquerades a valid user in the communication network to obtain access to the victim's confidential data.	• Email of HI business compromise refers to impersonation frauds in which a victim is misled into completing a cash transfer.

A. SUPERVISED LEARNING

Supervised learning is well known for fraud detection. It compares the newly arrived claims with the pre-trained models. In this method, the model is trained using data associated with the labels. The trained model can predict the newly arrived HIC data's class. The model can be built using a training dataset. The genuine and fraudulent claims could be the labels of class in the system to detect Medicare frauds [58]. If a claim follows a similar label, it can be classified as genuine, otherwise fraudulent.

Kose *et al.* [26] developed a framework that detects the HI abuse cases independently from the actors and commodities. They defined actors as the insured individuals, physicians, or institutional healthcare service providers and commodities as medication or healthcare services. They used the ZeroR classification supervised learning technique for fraud detection. They also utilized proactive and reactive analysis. The framework includes a visualization tool that significantly reduces the time requirements for the users during the fact-finding process after the RFID Suite alerted the user of risky claims. Then, Richard *et al.* [59] built and tested the anomaly detection model to flag outliers using publicly available 2013 and 2014 insurance data from the centre of Medicare and Medicaid Services. The testing is performed in the anomaly detection model to detect potentially fraudulent activities by predicting the medicare provider's specialty according to the number of procedures performed. Suppose a physician is expected to work in a different medical profession; they

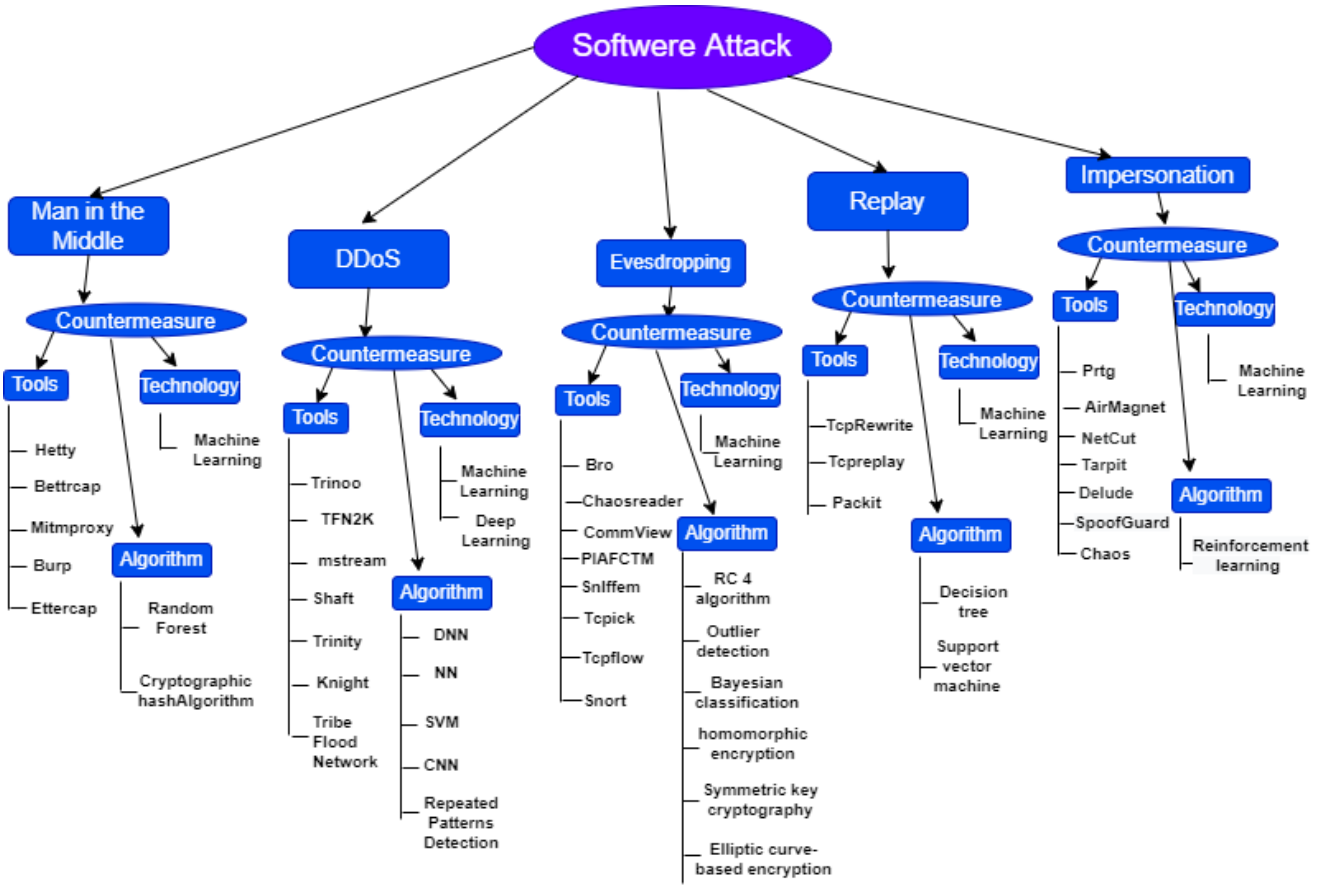

FIGURE 3: Possible software attack on HI and countermeasures.

likely have many one-of-a-kind patients or are engaging in potentially fraudulent activity. The strategies such as feature selection, a healthcare speciality, and grouping are utilized for fraud identification in healthcare insurance sector specialities. Later, the authors in Cassimiro [60] proposed a few experiments to examine the class imbalance impact of HIC fraud detection systems. This experiment predicts the performance of supervised techniques such as Repeated Incremental Pruning to Produce Error Reduction (RIPPER), SVM, random forest, and naïve Bayes due to the imbalanced class. They also used the recovery method Synthetic Minority Over-Sampling Technique (SMOTE) meta Cost and random oversampling to reduce the HI claims fraud. Later, Herland *et al.* [21] presented various methods for generating one dataset from three medicare CMS (HI claim) datasets. After that, data processing is performed and a map provider fraud label from the LEIE dataset to identify fraudulent behavior of physicians. For medical insurance fraud detection, supervised learning techniques like Random Forest, gradient boosted trees, and logistic regression was utilized for training three CMS medicare dataset and a combined dataset.

Then, Pandey *et al.* [61] developed a framework for identifying forged HIC. A scoring model created a fraud indicator to recognize forged claims for this goal. This fraud indicator classified the forged and non-forged claims. Linear regression is used to validate this scoring model. Sowah *et al.* [62] developed a decision-making method for detecting the HI fraud. SVM based on genetic algorithms, is utilized to classify fraudulent insurance claims from the National HI Scheme dataset. Later, Ilango *et al.* [63] proposed a framework that detects the HI fraud with faster learning from CMS medicare dataset using multi-Layer perceptron, a feed-forward neural network with genetic algorithm and also solves classification imbalance problem. Then, Yang *et al.* [64] proposed a novel light gradient boosting machine (LightGBM) mining algorithm for HI fraud detection. This algorithm eliminates the negative influence of the imbalance of positive and negative samples in training set by automatically selecting the hard examples and balancing the ratio between fraud samples and nonfraud samples. Table 5 shows the comparative analysis of various state-of-the-art supervised learning techniques for HIC fraud detection.

B. UNSUPERVISED LEARNING

In an unsupervised learning method, the model is trained using data without any associated label. Unsupervised learning can detect new and existing types of HI frauds as they aren't limited to fraud patterns with predefined class labels [58].

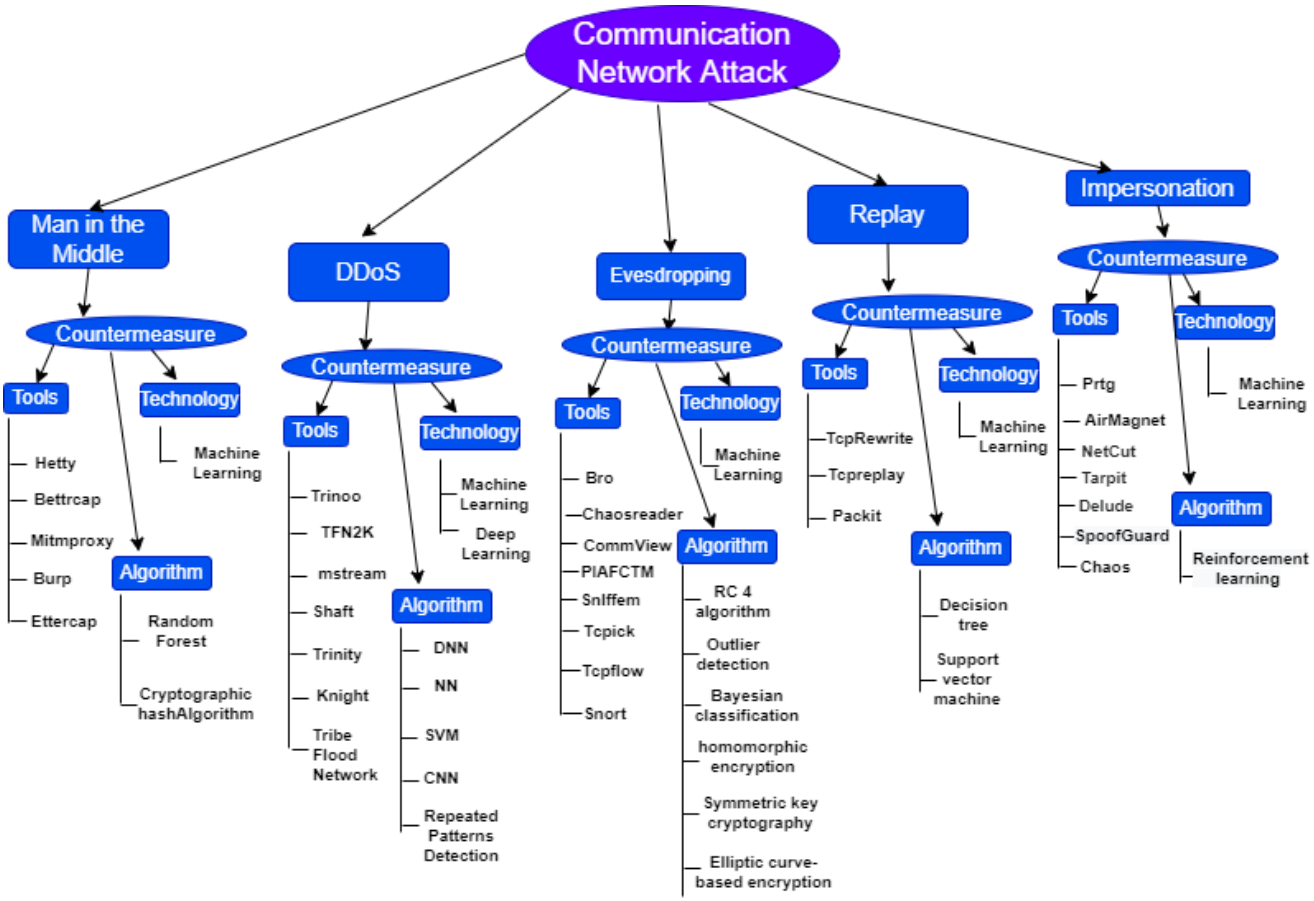

FIGURE 4: Possible communication network Attack on HI and countermeasures.

Verma *et al.* [66] proposed an approach for identifying the frequent fraud patterns from the HI database using rule mining which analyzed HIC fraudulent patterns according to period and disease. Period-based claim fraud outliers are identified using decision rules based on statistics and k-means clustering. In contrast, disease-based fraud outliers are identified using association rule mining and gaussian distribution. These outliers represent fraud insurance claims in the database. Then, Anbarasi *et al.* [67] proposed a framework that predicts the suspicious behavior of the HI fraud using probabilistic outlier detection. To reduce the time required for the fact-finding process, proactive and retrospective analysis are combined. Later, the authors in [68] proposed a self-organizing feature map neural network-based HI fraud detection model with Principal Component Analysis (PCA), which considers all the valuable features from the patient's claim data and improves the accuracy of the classification. PCA is utilized for dimension reduction of data. In this study, a few influence variables are selected to detect HI fraud. Table 6 shows the comparative analysis of existing unsupervised learning techniques used for HIC fraud detection.

C. HYBRID APPROACH

Some studies in the literature suggest a hybrid strategy of supervised and unsupervised learning methods for detecting HI frauds. This hybrid approach of ML technology provides efficient results in HI fraud identification. For example, Rawate *et al.* [58] proposed a supervised and unsupervised learning hybrid strategy for detecting fraud in the HI sector. This approach utilized the advantage of both classification and clustering techniques. Then, Kareem *et al* [69]. presented a method for recognizing the forged HIC by analyzing the correlations or associations between attributes on the HIC documents. They utilized an SVM supervised learning technique. This approach uses an unsupervised learning technique evolving clustering and association rule mining due to dynamic HI document data. Later, Jiang *et al.* [70] proposed a parallel framework to deal with the class imbalance and heterogeneous data of HI detection systems. This framework is cost-sensitive to deal with class imbalance. Deep learning algorithms such as Deep Neural Network (DNN), bidirectional long short-term memory (BiLSTM), Parallel Neural Network (PNN), and simulated neural network (SNN) are used to compare this framework's performance. Later, Zhou *et al.* [71] proposed a hybrid strategy for recognized forged repayment of the healthcare

TABLE 5: Comparative analysis of various state-of-the-art supervised learning techniques for HIC fraud detection.

Author	Year	Objective	Pros	Cons	Machine Learning Technique
Kose et al. [26]	2015	To develop a framework for detecting HI abusive claims from the actor and commodities. Actor refers to insurance subscribers, physicians, corporate healthcare service providers, and commodity refers to medication or healthcare services.	Even with the incomplete data, predicted the risky claim	The network between actors is not considered in the analysis process	Proactive and retrospective analysis, Zero R classification
Richard et al. [65]	2016	To develop a predictive model which detects fraudulent behavior in HIC based on classification of physician specialty	Effective prediction	Similar procedure perform by various type of physicians is not considering during fraud detection	Multinomial naive bayes, Logistic regression
Cassimiro et al. [60]	2017	To analyze the loss of HIC fraud prediction performance in brazillian healthcare due to class imbalance and present recovered method	Give comparison of various ML-based recovery methods for class imbalance issue	Availability of data set is limited, The small training data set	RIPPER, SVM, Random forest, Naive bayes
Herland et al. [21]	2018	To generate one dataset from three Medicare CMS datasets, perform data processing on it and perform mapping of provider fraud label for identifying fraudulent behavior of physician in HIC	Performance of HIC fraud detection from the combined generated dataset is best.	Class imbalance problem is not addressed during the exploratory analysis of medicare fraud	Logistic regression, Gradient boosted trees, Random forest
Pandey et al. [61]	2018	To build a scoring model which classifies fraud indicators of HIC and determines the best ML technique for HIC fraud detection	More accurate Result of fraud prediction and scoring for any new HIC data.	The model is unable to handle dynamic data.	Logistic regression, Decision tree, and Neural networks
Sowah et al. [62]	2019	To develop a model for HI fraud detection using genetic SVM.	Time taken for process claims is reduced due to enhancing Accuracy of classification.	High computation time	Genetic support vector machine, Decision tree, and naive Bayes
Ilango et al. [63]	2020	To develop a framework for discovering the more number of fraud patterns in HIC using by comparative analysis of various supervised learning techniques	More accurate and time efficient	Not given a detailed analysis of model training time	Multi-layer perceptron with a genetic algorithm, Feed-forward Neural Network, Logistic regression, GaussianNB classification, Decision tree classifier
Yang et al. [64]	2021	To develop bootstrapping HIC fraud detection model by choosing fraud samples and removing a legitimate sample for balance between fraud sample and legitimate sample	The model takes less time for training.	Small dataset is taken for model training.	Light gradient boosting machine based hard example mining algorithm(LHEM)

insurance sector. Local outlier factors and clustering are combined in this hybrid approach. This approach provides better results for newly arrived insurance claim data. Table 7 shows the comparative analysis of existing Hybrid approaches for HIC fraud detection.

D. HIC FRAUD DETECTION FROM GROUP OF PEOPLE

As discussed in Section II-B various types of HIC frauds, a group of people are involved with it. Some researchers are analyzing HI fraud detection methods from a group of people. For example, Wanga *et al.* [72] presented an approach for detecting dentist fraudulent claims based on the trustworthiness score of the dentist, which is evaluated from the social network of patient and dentist. In this approach, the suspicious link between two or many dentists is analyzed based on the social network. Later, Figueredo *et al.* [73] developed a HI fraud detection model which detects the fraud from insurance subscriber's claim data. Fraud can identify based on mutual referrals among groups of physicians. Social networking techniques are used to find mutual referrals among groups of physicians. Then, Sun *et al.* [74] developed a method to detect joint fraud based on fraudulent group mining in the HI sector. In this method, mining is evaluated using a similarity adjacency graph and classifies abnormal

TABLE 6: Comparative analysis of existing unsupervised learning techniques used for HIC fraud detection.

Author	Year	Objective	Pros	Cons	Machine learning algorithm
Verma et al. [66]	2017	To identify HI fraud patterns periodically and disease wise	It contains a common strategy to generate a prediction of identifying medicare insurance fraud in many healthcare specialties, which is applied flexibly at scale.	This study has no effective strategies for fraud prevention and does not identify newly emerging fraud	K-means clustering, association rule mining, elbow test, and outlier detection
Anbarasi et al. [67]	2017	To identify abnormal fraud patterns in the HIC system using outlier detection and proactive, reactive analytics integration	The amount of time taken by the process which finds facts is reduced	Not given information about a tool for identifying nature of HIC fraud occurrences	Outlier detection and analytic hierarchical processing's pairwise comparison technique
Cao *et al.* [68]	2019	To extract features from the medical claim database which improve Accuracy of HI fraud classification model using self-organized feature map neural network.	Overall time complexity lowered, The cost of analysis decreased, The overall accuracy of classification has enhanced	The fraud indicator of healthcare insurance data was not precise, and it impacted the model's detection ability.	Self-organizing feature map and PCA

TABLE 7: Comparative analysis of existing Hybrid approaches for HIC fraud detection.

Author	Year	Objective	Pros	Cons	Machine learning Technique
Rawte et al. [58]	2015	To develop hybrid model using SVM supervised learning (SVM) and unsupervised learning (ECM) for identify duplicate HIC	Scalability	Accuracy information is missing	Evolving Clustering Method (ECM) and SVM
Kareem et al. [69]	2017	To identifying correlation or link between features of HIC record to detect the forged HI claims	Scalability	Only the pre-processing forged HIC detection was addressed	ECM, association rule mining, and SVM
Jiang et al. [70]	2018	To develop a parallel framework for fraud detection in HIC operation from real-world sequence and descriptive data by comparative analysis of various deep learning approaches	Processed heterogeneous data and classify imbalance data	-	DNN, BiLSTM, PNN, and SNN
Zhou et al. [71]	2020	To propose a method for detecting fraudulent payment in HIC reimbursement the process from real HIC data using local outlier factor and clustering	Scalability	Fraud detection analyzed only from two diseases wise	Outlier detection and clustering

groups according to the similarity adjacency graph. Table 8 shows the comparative analysis of existing HIC Fraud detection systems based on group of people.

E. BLOCKCHAIN BASED HIC FRAUD PREVENTION

Liu *et al.* [75] proposed a blockchain-based healthcare insurance system, i.e., a cloud-based architecture. This system gives medical insurance an anti-fraud service, which determines whether a patient's medical reimbursement request is genuine and fits the policy's requirements. Later, Gera *et al.* [76] developed a framework and consensus mechanism of three peers, which are police, insurance firm staff, and agents, for solving the security issues in the HI transactions. Every transaction can be stored as a chain of blocks and cryptographically signed in the IBM blockchain platform to prevent HIC transactions from fraud. Later, Saldamli *et al.* [77] provides a blockchain-based solution for HI fraud committed by the patient or any malicious entity. In this solution, transactions associated with the patient's health claim are tracked using blockchain. If a claim transaction resembles a past claim transaction, then the patient's insurance application request needs to be declined in the network. Then, Ismail *et al.* [78] proposed a taxonomy and HI fraud detection framework using blockchain and examined results according to the time taken by execution and data transmitted when the number of HIC is increased. In this framework, 12 different fraud scenarios are detected, and detection was performed on patients, doctors, pharmaceutical firms, and doctors. Later, Baker *et al.* [79] developed a consortium blockchain-based distributed architecture and consensus mechanism to restrain duplication in the HIC system. They analyze the time required to validate the transactions (validation time), the time needed to upload transactions into the transaction pool (upload time), the time required for inserting blocks into the chain, security, and privacy of data. Table 9 shows a comparative analysis of existing blockchain-based systems for HIC fraud prevention.

TABLE 8: Comparative analysis of existing HIC Fraud detection systems based on group of people.

Author	Year	Objective	Pros	Cons	Machine learning algorithm
Wanga et al. [72]	2016	To detect fraudulent dentist claims based on the trustworthiness score of the dentist, which is evaluated from the social network of patient and dentist.	HIC application evaluation is quick, Reduce Claim reviewer's exhaustive workload and improve their review accuracy,	Some fraudulent dentists cannot be identified because their patients' loyalty prohibits them from building social relationships between them.	Zero R classification
Figueredo et al. [73]	2018	To disclose common reference between physicians and identify fraud pattern of physician's practice from HIC data.	Accuracy of the model is excellent.	Clinical test or surgeries was not analyzed in the fraud detection model.	Page rank algorithm, Social network analysis
Sun et al. [74]	2019	TO develop a method for detecting joint fraud in HIC based on fraudulent group mining. This mining is evaluated using a similarity adjacency graph and classifies abnormal Groups according to a similarity adjacency graph.	More precise and reduce calculation	Joint fraud detection performed on a limited number of people's record	L-SVM, Abnormal grouping, and Density-based incremental local outlier factor

TABLE 9: Comparative analysis of existing blockchain-based systems for HIC fraud prevention.

Author	Year	Objective	Pros	Cons	Protocol/Technique
Liu et al. [75]	2019	To develop blockchain-based HIC system to prevent forged data, third-party accidental fraud risk, repayment of faulty electronic bill	Provide better data privacy preservation, adaptability	Not given implementation results	GRPC and Gossip protocol
Gera et al. [76]	2020	To developed framework for preventing HI fraud transaction using IBM platform of blockchain and developed consensus involved three peer police, agent and HI firm staff	Security, Non-repudiation, Integrity	Scalability	Proof-of-Work
Saldamli et al. [77]	2020	To track all patient's HIC transactions and design a HIC record secure sharing solution using blockchain, which helps to detect fraud in HIC	Easy HIC fraud detection, Efficiency of middleware to scan and communication of HIC record is good, time taken for data processing in blockchain was less	Not stable version of database	-
Ismail et al. [78]	2021	Developed a blockchain-based taxonomy for HI fraud detection considering 12 different fraud scenarios.	Maintain performance when healthcare insurance branches and claims transactions grow.	Interoperability issue	NS 3
Baker et al. [79]	2021	To develop blockchain-based distributed architecture and consensus mechanism to prohibit duplication in medicare insurance claims.	Frameworks are run very fast and increase the security and privacy of data	Scalability	First-in-first-out, Proof-of-work, and Proof-of-stake.

F. AI AND BLOCKCHAIN-BASED HIC FRAUD DETECTION

Integration of blockchain and ML for HIC fraud detection solved many issues of HIC, such as security, privacy, and data interoperability. Blockchain provides secure HIC data storage and tamper-proof HIC data transfer in which insurance subscribers consent to share their records with the parties involved with HIC. ML can use this HIC data to identify a fraudulent pattern and generate correct predictions about fraudulent claims. Two studies found in the literature show the integration of ML and blockchain for HIC fraud detection. Dhieb *et al.* [80] proposed a framework for HIC fraud detection and risk evaluation in which ML techniques are performed on data stored in the blockchain. Data cleaning is performed for better performance results of the fraud detection analysis. They also performed a comparative analysis of fraud indicators using different ML techniques such as SVM, XGBoost, fast decision tree, Stochastic Gradient Descent, Naïve Bayes, and Nearest Neighbour. Blockchain network is created using hyperledger fabric, and for integration between blockchain and ML, they used Rest API. Zhang *et al.* [81] developed a framework for identifying HIC fraud based on bidirectional encoder representations from transformers (BERT-LE) deep learning techniques using two real hospital datasets. BeRT-LE technique was used to classify the reasonability of the

International Classification of Diseases-10 (ICD-10) code, which is attached with the E-health report. The reasonability of the ICD code is evaluated from the inpatient's description of their sickness, which they called a chief complaint. They also proposed a health record storage and management method based on the consortium blockchain for data security, reliability, immutability, traceability, and non-repudiation. Table 10 shows a comparative analysis of existing AI and Blockchain-based HIC fraud detection systems.

V. THE PROPOSED SYSTEM

In this section, we propose a layered intelligent HI fraud detection system. The number of layers are four which are named as user layer, data generation layer, data analytical layer, and blockchain layer. FIGURE 5 shows the proposed four-layer architecture. The detailed description of each layer is described in subsequent subsections.

A. USER LAYER

There are three primary users involved in the proposed HIC fraud detection architecture, such as insurance provider, healthcare service provider, and insurance subscriber. Insurance providers include private and government insurance firms. Healthcare service providers include hospitals, pharmacies, medical labs, ambulance services, and doctors. The insurance subscriber is a patient, who request an HI claim. This layer is connected to a blockchain network in which each activity of all the users are recorded. Apart from the blockchain, this layer is also connected to the data generation layer.

B. DATA GENERATION LAYER

There are two sources for the HIC data collection, which are offline and online processes. Offline data sources include paper-based billing forms, drug prescriptions, Paper based HIC, medical guideline books. Offline data sources such as billing forms, drug prescriptions, and medical guideline books are collected from hospitals (healthcare service provider), ambulance service providers, medical laboratories, pharmacies. Paper based HIC is collected from the HI government sector, or private firms (HI provider). Offline collected HIC data are digitized through internet. Online data sources include publicly available research-based datasets, EHR, HIC data in an electronic form, and social media data which are available through internet. EHR are collected from healthcare service provider and social media data are collected from insurance subscriber (patient). After data collection, the data is generated according to the HIPPA act. Data generated from this layer will share and control through smart contract condition for security purposes because this data contains payment-related sensitive information of insurance subscribers and providers.

C. DATA ANALYTICAL LAYER

HIC data have sparsity, heteroscedasticity, multicollinearity, and missing values. The best way to deal with such data is to use data preprocessing and apply robust ML approaches. In the data analytics layer, data preprocessing is involved with data cleaning, data integration, data transformation, and data reduction. In the data cleaning process, all redundant, incomplete, and unnecessary HIC data are eliminated. HIC data generated from all sources are combined in the data integration process. In the data transformation process, the structure of the HIC data is modified through various techniques such as smoothing, aggregation, and normalization. The data reduction process decreases the HIC data storage, making HIC data analysis easier while producing similar results. After data preprocessing, data is ready for data analytics, which can be performed using machine learning algorithms according to fraud indicators. Various ML algorithms are used to determine if a claim is legitimate or fraudulent. This prediction result is integrated with smart contract through blockchain layer.

D. BLOCKCHAIN LAYER

The blockchain layer provides a secure environment for all payment and authentication-related transactions. Each layer of architecture can communicate with the blockchain layer through a smart contract. Smart contracts are self-executing predefined rules for authentication and insurance coverage payment. In a smart contract, consensus algorithms like PoW, proof of burn (PoB), or proof of stake (PoS) are utilized to authenticate every transaction and input. A smart contract is integrated with InterPlanetary File System (IPFS) because it minimizes the overall data storage and cost for fast data retrieval. Blockchain is a decentralized network of peers that communicate using a smart contract that eliminates the need for a third party. In FIGURE 5, blockchain layer shows how all HIC transactions are performed. Cryptography algorithms and the immutable qualities of the blockchain enhance the security of the HIC data. When HIC data is copied into the blockchain, it is impossible to change it.

VI. CASE STUDY ON HI FRAUD DETECTION

Nowadays, people have become more conscious of their healthy lifestyle choices due to the Covid-19 pandemic. They have begun to live healthier lifestyles using advanced technologies such as wearables and smartphone healthcare apps. A wearable device is a smart electronic device that can be worn as jewellery, embedded in clothing or body [84]. Smartphone healthcare apps provide healthcare-related services. Data captured from healthcare apps and wearable devices can be used for clinical research. Some HI provider firms use wearable devices and healthcare apps data containing blood sugar tracker data, fitness tracker data, pedometer data, and telehealth data [85] [86]. They use wearable devices as part of their marketing scheme to connect with insurance subscribers. Under certain conditions, HI provider firms offer rewards or discounts on the premium price and free wearables as part of their marketing scheme. Conditions are insurance subscribers use a wearable, give consent about wearable tracker data sharing, and show

TABLE 10: Comparative analysis of existing AI and Blockchain-based HIC fraud detection systems.

Author	Year	Objective	Pros	Cons	Platform/Algorithm
Dhieb et al. [80]	2020	To develop a framework for HIC fraud detection and risk evaluation using blockchain and machine learning	Easily integrate module into the system, scalable	Authorized party check fraud claim generated from ML module manually	Hyperledger fabric, XGBoost, Very fast decision tree, Stochastic gradient descent, Naïve bayes, Nearest neighbour, SVM.
Zhang et al. [81]	2021	Proposed a framework for detection of HIC fraud using text classification, and HIC data are stored on consortium blockchain which prevents from intrusion.	Reduce HI auditor workload, Secure, traceable, and non-repudiation data storage	Result of prediction was biased, Not consider outpatient scenario during training	DPCNN, Text-RNN, HAN, FastTex, LEAM, BERT, and BERT-LE

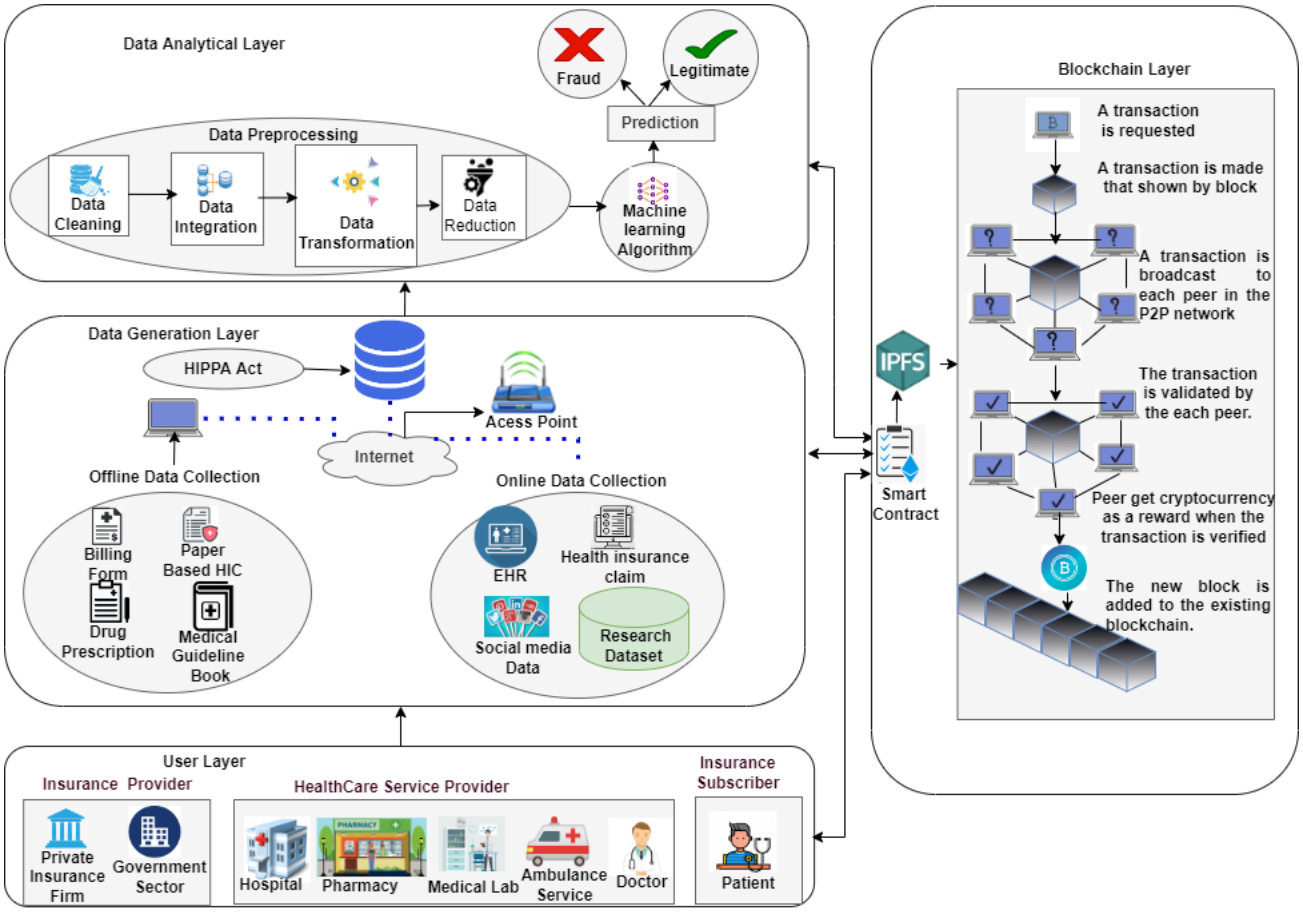

FIGURE 5: Architecture of HI Fraud Detection [82], [83]

healthy behaviors such as walking a certain number of steps each week, count of calories, maintaining a healthy heart rate through wearables [84].

One of the US-based insurance firm launched a platform in 2016 that motivate and reward the insurance subscriber to participate in healthy activities through wearable devices. According to the tracked data from wearable devices, that firm provides challenges and competitions to insurance subscribers. Insurance subscribers are rewarded with fitness equipment, gift cards, and reduced HI premium rates [87]. Another US partnered firm distribute Misfit fitness trackers to all the insurance subscribers for free. This firm is one of the first insurance provider firms who evaluate the insurance premium rate of insurance subscribers through wearables data. They linked insurance subscribers' biometric data with their HI account [87]. Then France-based HI firm encourage insurance subscribers to share their health-related personal data from a fitness device. Insurance subscribers can earn fit points by sharing their fitness device data. The firm uses this data for its insurance underwriting process [87]. Then Autralia-based insurance provider firm, formed a partnership to integrate insurance subscriber's data from various wearable fitness trackers to their accounts. This data is used for insurance risk assessment, and pricing of

premiums [87].

As discussed above, the insurance firm uses wearable device data only to determine the pricing of premiums and insurance risk assessment. They are not using it for fraud detection in HIC. The HI fraudsters become smart day by day and find new ways to do fraud. So, there is a strict requirement for any insurance firm to find a new and improved futuristic way to combat fraud using wearable device data. Hence, we present a novel case study to fulfil the above requirement of insurance firms. FIGURE 6 shows the flow of the case study, which solves the security and privacy issues of HI.

In FIGURE 6, there are various wearable devices worn by insurance subscribers, such as smart glasses, blood pressure sensors, fitness bands, smart shoes, etc. The insurance firm may provide wearable for marketing and fraud detection in HIC. Wearable wear by insurance subscribers is connected with insurance subscribers' smartphones through WiFi or Bluetooth. The data generated from wearables are stored in IPFS, which is in encrypted form and digitally signed. Insurance providers and healthcare service provider data are encrypted, digitally signed, and uploaded to the fraud detection system. After data generation, data preprocessing includes data cleaning, integration, transformation, and reduction method. Before data preprocessing, insurance subscribers distribute consent about sharing personal healthcare data with a predefined Ethereum smart contract. The insurance provider firm provides the public key to insurance subscribers and healthcare providers. The insurance firm will reward cryptocurrency to insurance subscribers if insurance subscribers give consent about sharing personal healthcare data. An ML algorithm is applied to preprocessed data that predict that HIC is fraud or legitimate. Blockchain network consensus enables automated claim coverage payment to insurance subscribers if a claim is legitimate. But, if the claim is fraudulent, the subscriber would pay the penalty.

A. REAL-TIME ANALYTICS

Earlier, the HIC management system relies on insurance subscriber reporting. Analysis and auditing of the HIC are done after the hospitalization emergency occurs. At the same time, HIC management uses wearable analysis for real-time events such as cardiovascular events. It also increases insurance subscribers' engagement with the insurance provider in real-time and frequently. The location of the insurance subscribers will be traced using the wearable device. Location data and the real-time health status of insurance subscribers can observe what happened before, during, and after a hospitalization emergency. Insurance firms can use this data to detect pain levels, counteract medication-seeking behavior, and discover inconsistent behavior with a HIC. Earlier, real-time data is not available to the HIC fraud detection system.

B. SECURITY AND PRIVACY

Data use in the HIC is sensitive as it contains payment-related and insurance subscriber's personal health-related sensitive information. Insurance subscriber's data can be exchanged among all parties involved with HIC fraud detection without revealing the insurance subscriber's identities because the identities of insurance subscribers in a blockchain are pseudonymized using cryptographic keys. The proposed system uses encryption, decryption, and digital sign to secure the HIC system from various attacks and privacy breaches.

VII. OPEN ISSUES AND RESEARCH CHALLENGES

This section highlights research challenges and open issues of blockchain and AI-empowered healthcare insurance fraud detection.

A. CLASS IMBALANCE

Class imbalance is the vital challenge that reduces the performance of the HIC fraud classifier built based on ML models. The majority class in the imbalance classification predictive modelling problem has more records than the minority (fraud) class [19]. The minority class is more challenging to predict because it contains fewer HIC records. Thus, a detailed and advanced solution is needed to handle the class imbalance issue.

B. STANDARDIZATION OF DATA EXCHANGE RULE

There is no universal rule or regulation for the cross-border exchange of HIC data. The government privacy rule for HIC data exchange differs from one country to another [13]. The advantage of blockchain and AI-enabled HIC fraud detection system data sharing may be restricted by the cross-border exchange of HIC data where different jurisdictions are involved. So, there is a need to do further research on the standardization of cross-border HIC data retrieval, storage rules, and regulations.

C. REAL TIME DATA-SET ACCESSIBILITY

The public and private healthcare sectors should provide more opportunities for researchers to access real-world HIC data. The availability of real-world HIC data sets has been restricted due to the legal, privacy, and competitive concerns [88]. Future research with open source HIC real-world datasets is essential, allowing other researchers to validate their results. Researchers should collaborate with the healthcare insurance sector to access the proposed solutions using real-world healthcare insurance data.

D. HANDLING GROWING NUMBER OF USERS

One of the challenging tasks is to handle the growing number of users in AI and blockchain-enabled HIC fraud detection framework. Due to the population growth, the need for the healthcare insurance sector is increasing for the well-being of patients. As the number of insurance subscribers on the framework grows, AI and blockchain-

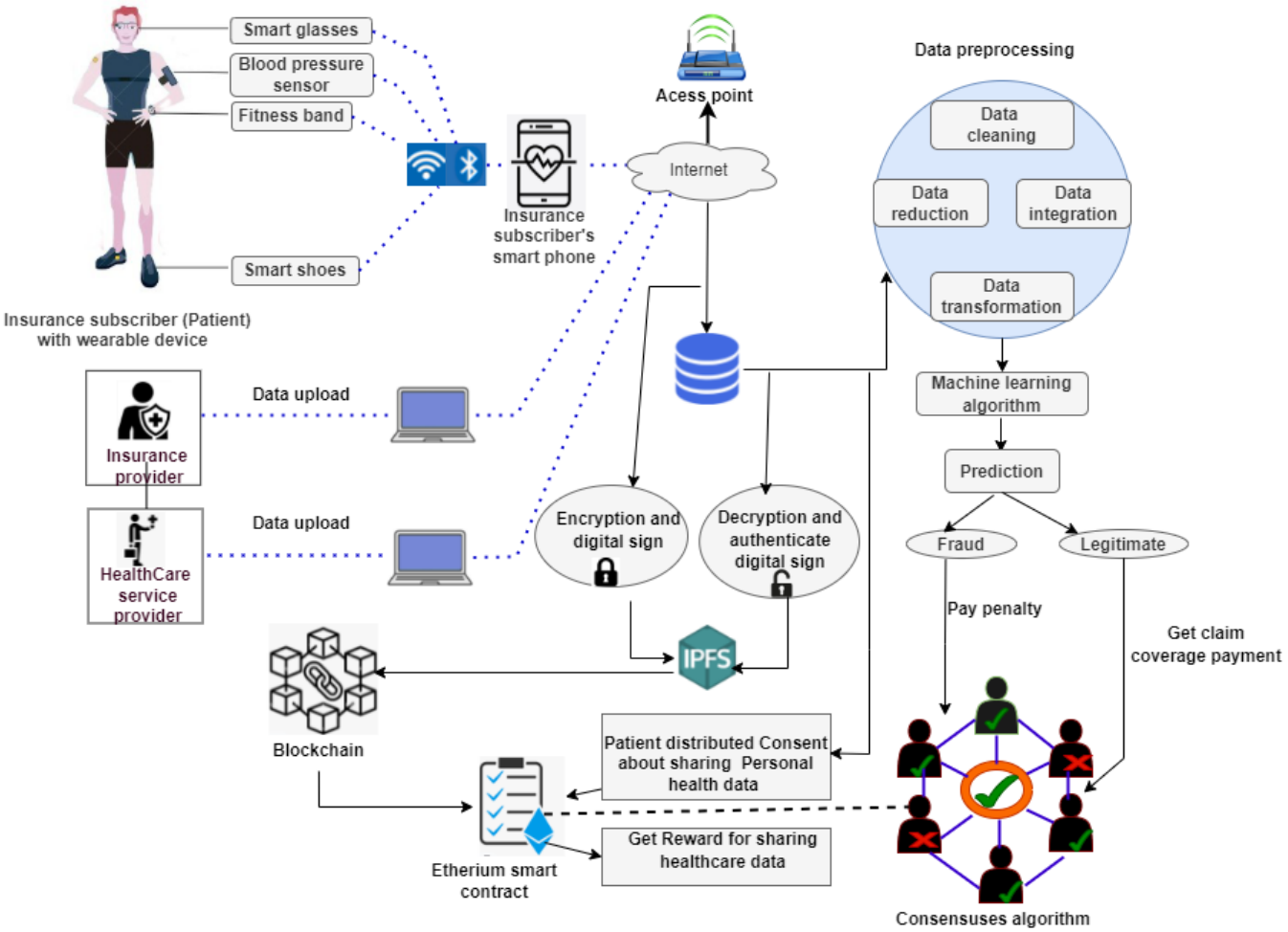

FIGURE 6: Case study : HI Fraud Detection from wearable device.

enabled frameworks become more challenging to implement in real-time scenarios.

E. AMALGAMATION WITH OTHER TECHNOLOGY

HI firms have a scope to amalgamate with wearables and sensor technologies used in the healthcare for the additional functionality of the business operations [89]. HI providers can utilize wearable devices to collect real-time data of insurance subscribers, to analyze their premium rates and HIC fraud detection. So, future research work should focus on the wearable and IoT-based healthcare devices data availability and data analytics for blockchain and AI-enabled HIC fraud detection system.

F. LACK OF PROFESSIONAL TALENT

Blockchain and AI have become one of the most rapidly emerging technologies in the present era. Developing blockchain and AI-enabled HIC fraud detection applications require professionally skilled people who know the complex AI, cryptography, and advanced mathematics algorithms. Currently, only some researchers and professionals are working on ensuring blockchain and AI-based schemes are robust and adaptive. So, adapting the AI and blockchain in HIC fraud detection will be a challenging task due to the lack of professional talent.

G. WEAKNESS OF SMART CONTRACT DEVELOPMENT

Smart contracts are executable codes designed in solidity and deployed on Ethereum Virtual Machines to regulate action according to the agreement. In the development phase of the smart contract, there is a possibility of a software bug. There is a requirement to handle the software bug before the smart contracts are deployed in the blockchain network, as they can't be altered after the deployment. So, research solutions should be improved to deploy the smart contract efficiently and reliably.

VIII. CONCLUSION

We presented a systematic survey on HI fraud detection and associated security issues in HIC. Firstly, we highlight the background context of HI in which the revolution of HI, types of HI fraud are covered. The various security issue of HI, possible attacks on HIC, countermeasure tools,

technologies, and algorithms that protect against the possible attacks are discussed. This paper described a category-wise taxonomy of HIC fraud detection. We proposed a four-layer intelligent HI fraud detection system architecture. Further, this paper has introduced a futuristic approach for the proposed HIC fraud detection system using a wearable device. Also, we have listed various research challenges and the open issue associated with the blockchain and AI-based proposed system during its real-time deployment.

REFERENCES

[1] H. insurance, "Health insurance in india." https://en.wikipedia.org/wiki/Health_insurance_in_India. Accessed: 2021.

[2] A. Sheshasaayee and S. S. Thomas, "A purview of the impact of supervised learning methodologies on health insurance fraud detection," in Information Systems Design and Intelligent Applications, pp. 978–984, Springer, 2018.

[3] H. K. Patil and R. Seshadri, "Big data security and privacy issues in healthcare," in 2014 IEEE international congress on big data, pp. 762–765, IEEE, 2014.

[4] M. Ojha and K. Mathur, "Proposed application of big data analytics in healthcare at maharaja yeshwantrao hospital," in 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1–7, IEEE, 2016.

[5] M. Eling and M. Lehmann, "The impact of digitalization on the insurance value chain and the insurability of risks," The Geneva papers on risk and insurance-issues and practice, vol. 43, no. 3, pp. 359–396, 2018.

[6] R. Dutt, "The impact of artificial intelligence on healthcare insurances," in Artificial Intelligence in Healthcare, pp. 271–293, Elsevier, 2020.

[7] P. Kumar and H.-J. Lee, "Security issues in healthcare applications using wireless medical sensor networks: A survey," sensors, vol. 12, no. 1, pp. 55–91, 2012.

[8] M. Li, W. Lou, and K. Ren, "Data security and privacy in wireless body area networks," IEEE Wireless communications, vol. 17, no. 1, pp. 51–58, 2010.

[9] J. Heurix, M. Karlinger, and T. Neubauer, "Pseudonymization with metadata encryption for privacy-preserving searchable documents," in 2012 45th Hawaii International Conference on System Sciences, pp. 3011–3020, IEEE, 2012.

[10] K. M. Kumar, T. S, and S. Swarnalatha, "Effective implementation of data segregation amp; extraction using big data in e - health insurance as a service," in 2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 01, pp. 1–5, 2016.

[11] D. Ulybyshev, C. Bare, K. Bellisario, V. Kholodilo, B. Northern, A. Solanki, and T. O'Donnell, "Protecting electronic health records in transit and at rest," in 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pp. 449–452, IEEE, 2020.

[12] "The benefits and threats of blockchain technology in healthcare: A scoping review," International Journal of Medical Informatics, vol. 142, p. 104246, 2020.

[13] C. C. Agbo, Q. H. Mahmoud, and J. M. Eklund, "Blockchain technology in healthcare: A systematic review," Healthcare, vol. 7, no. 2, 2019.

[14] E. A. Duman and Ş. Sağıroğlu, "Heath care fraud detection methods and new approaches," in 2017 International Conference on Computer Science and Engineering (UBMK), pp. 839–844, IEEE, 2017.

[15] R. Bauder and T. Khoshgoftaar, "A survey of medicare data processing and integration for fraud detection," in 2018 IEEE international conference on information reuse and integration (IRI), pp. 9–14, IEEE, 2018.

[16] T. Ekin, F. Ieva, F. Ruggeri, and R. Soyer, "Statistical medical fraud assessment: exposition to an emerging field," International Statistical Review, vol. 86, no. 3, pp. 379–402, 2018.

[17] D. Ankrah, J. Hallas, J. Odei, F. Asenso-Boadi, L. Dsane-Selby, and M. Donneyong, "A review of the ghana national health insurance scheme claims database: possibilities and limits for drug utilization research," Basic & clinical pharmacology & toxicology, vol. 124, no. 1, pp. 18–27, 2019.

[18] Z. X. Chen, L. Hohmann, B. Banjara, Y. Zhao, K. Diggs, and S. C. Westrick, "Recommendations to protect patients and health care practices from medicare and medicaid fraud," Journal of the American Pharmacists Association, vol. 60, no. 6, pp. e60–e65, 2020.

[19] A. J. Mary and S. A. Claret, "Imbalanced classification problems: Systematic study and challenges in healthcare insurance fraud detection," in 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1049–1055, IEEE, 2021.

[20] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: a comprehensive review from 2009 to 2019," Computer Science Review, vol. 40, p. 100402, 2021.

[21] M. Herland, T. M. Khoshgoftaar, and R. A. Bauder, "Big data fraud detection using multiple medicare data sources," Journal of Big Data, vol. 5, no. 1, pp. 1–21, 2018.

[22] M. Ahmed and M. Ahamad, "Combating abuse of health data in the age of ehealth exchange," in 2014 IEEE International Conference on Healthcare Informatics, pp. 109–118, 2014.

[23] Healthcare, "Healthcare in india." https://en.wikipedia.org/wiki/Healthcare_in_India. Accessed: 2021.

[24] P. M. Suraksha, "Pradhan mantri suraksha bima yojana." https://financialservices.gov.in/insurance-divisions/Government-Sponsored-Socially-Oriented-Insurance-Schemes/Pradhan-Mantri-Suraksha-Bima-Yojana(PMSBY). Accessed: 2021.

[25] P. M. garib kalyan, "pradhan mantri garib kalyan-package." https://www.india.gov.in/spotlight/pradhan-mantri-garib-kalyan-package-pmgkp. Accessed: 2021.

[26] I. Kose, M. Gokturk, and K. Kilic, "An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance," Applied Soft Computing, vol. 36, p. 283–299, 08 2015.

[27] N. Rayan, "Framework for analysis and detection of fraud in health insurance," in 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 47–56, 2019.

[28] N. Rayan, "Framework for analysis and detection of fraud in health insurance," in 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 47–56, 2019.

[29] S. Khezr, M. Moniruzzaman, A. Yassine, and R. Benlamri, "Blockchain technology in healthcare: A comprehensive review and directions for future research," Applied sciences, vol. 9, no. 9, p. 1736, 2019.

[30] L. Settipalli and G. Gangadharan, "Healthcare fraud detection using primitive sub peer group analysis," Concurrency and Computation: Practice and Experience, p. e6275, 2021.

[31] C. Verma, V. Stoffová, Z. Illés, S. Tanwar, and N. Kumar, "Machine learning-based student's native place identification for real-time," IEEE Access, vol. 8, pp. 130840–130854, 2020.

[32] D. Reebadiya, T. Rathod, R. Gupta, S. Tanwar, and N. Kumar, "Blockchain-based secure and intelligent sensing for autonomous vehicles activity tracking beyond 5g networks," Peer-to-Peer Networking and Applications, 02 2021.

[33] M. H. Nasir, J. Arshad, M. M. Khan, M. Fatima, K. Salah, and R. Jayaraman, "Scalable blockchains—a systematic review," Future Generation Computer Systems, vol. 126, pp. 136–162, 2022.

[34] J. Shah, S. Agarwal, A. Shukla, S. Tanwar, S. Tyagi, and N. Kumar, "Blockchain-based scheme for the mobile number portability," Journal of Information Security and Applications, vol. 58, p. 102764, 2021.

[35] R. Kakkar, R. Gupta, S. Tanwar, and J. J. P. C. Rodrigues, "Coalition game and blockchain-based optimal data pricing scheme for ride sharing beyond 5g," IEEE Systems Journal, pp. 1–10, 2021.

[36] A. Kumari, R. Gupta, S. Tanwar, S. Tyagi, and N. Kumar, "When blockchain meets smart grid: Secure energy trading in demand response management," IEEE Network, vol. 34, no. 5, pp. 299–305, 2020.

[37] R. Saranya and A. Murugan, "A systematic review of enabling blockchain in healthcare system: Analysis, current status, challenges and future direction," Materials Today: Proceedings, 2021.

[38] issues facing patient privacy, "top 3 issues facing patient privacy." https://www.healthcareitnews.com/news/top-3-issues-facing-patient-privacy. Accessed: 2021.

[39] C. Thapa and S. Camtepe, "Precision health data: Requirements, challenges and existing techniques for data security and privacy," Computers in biology and medicine, vol. 129, p. 104130, 2021.

[40] atlantatech, "addressing the inadequacies of hipaa law and politics." https://www.atlantatech.news/analysis/addressing-the-inadequacies-of-hipaa-law-and-politics-involving-healthcare-it/. Accessed: 2021.

[41] A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, and I. E. Davidson, "Ransomware detection, avoidance, and mitigation scheme: A review and future directions," Sustainability, vol. 14, no. 1, 2022.

[42] S. H. Kok, A. Abdullah, N. Jhanjhi, and M. Supramaniam, "Prevention of crypto-ransomware using a pre-encryption detection algorithm," Computers, vol. 8, no. 4, 2019.
[43] G. Singh, D. Kant, U. Gangwar, and A. P. Singh, "Sql injection detection and correction using machine learning techniques," in Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1, pp. 435–442, Springer, 2015.
[44] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," Computers & Security, vol. 68, pp. 160–196, 2017.
[45] R. Raveendranath, V. Rajamani, A. J. Babu, and S. K. Datta, "Android malware attacks and countermeasures: Current and future directions," in 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 137–143, 2014.
[46] G. K. Sadasivam, C. Hota, and B. Anand, Honeynet Data Analysis and Distributed SSH Brute-Force Attacks, pp. 107–118. Singapore: Springer Singapore, 2018.
[47] S. Reddy and G. K. Shyam, "A machine learning based attack detection and mitigation using a secure saas framework," Journal of King Saud University - Computer and Information Sciences, 2020.
[48] R. Zagrouba and R. AlHajri, "Machine learning based attacks detection and countermeasures in iot," International Journal of Communication Networks and Information Security (IJCNIS), vol. 13, no. 2, 2021.
[49] M. S. Yousefpoor, E. Yousefpoor, H. Barati, A. Barati, A. Movaghar, and M. Hosseinzadeh, "Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: A comprehensive review," Journal of Network and Computer Applications, vol. 190, p. 103118, 2021.
[50] S. Tu, M. Waqas, S. U. Rehman, T. Mir, G. Abbas, Z. H. Abbas, Z. Halim, and I. Ahmad, "Reinforcement learning assisted impersonation attack detection in device-to-device communications," IEEE Transactions on Vehicular Technology, vol. 70, no. 2, pp. 1474–1479, 2021.
[51] N. Bruce, M. Sain, and H. J. Lee, "A support middleware solution for e-healthcare system security," in 16th International Conference on Advanced Communication Technology, pp. 44–47, 2014.
[52] "Ransomware: Recent advances, analysis, challenges and future research directions," Computers & Security, vol. 111, p. 102490, 2021.
[53] V. Akashe, R. L. Neupane, M. L. Alarcon, S. Wang, and P. Calyam, "Network-based active defense for securing cloud-based healthcare data processing pipelines," in 2021 International Conference on Computer Communications and Networks (ICCCN), pp. 1–9, 2021.
[54] S. A. Roseline and S. Geetha, "A comprehensive survey of tools and techniques mitigating computer and mobile malware attacks," Computers & Electrical Engineering, vol. 92, p. 107143, 2021.
[55] M. M. Najafabadi, T. M. Khoshgoftaar, C. Kemp, N. Seliya, and R. Zuech, "Machine learning for detecting brute force attacks at the network level," in 2014 IEEE International Conference on Bioinformatics and Bioengineering, pp. 379–385, IEEE, 2014.
[56] A. M. Amin and M. S. Mahamud, "An alternative approach of mitigating arp based man-in-the-middle attack using client site bash script," in 2019 6th International Conference on Electrical and Electronics Engineering (ICEEE), pp. 112–115, 2019.
[57] K. Xylogiannopoulos, P. Karampelas, and R. Alhajj, "Early ddos detection based on data mining techniques," in IFIP International Workshop on Information Security Theory and Practice, pp. 190–199, Springer, 2014.
[58] V. Rawte and G. Anuradha, "Fraud detection in health insurance using data mining techniques," in 2015 International Conference on Communication, Information & Computing Technology (ICCICT), pp. 1–5, IEEE, 2015.
[59] M. Herland, R. A. Bauder, and T. M. Khoshgoftaar, "Medical provider specialty predictions for the detection of anomalous medicare insurance claims," in 2017 IEEE International Conference on Information Reuse and Integration (IRI), pp. 579–588, 2017.
[60] J. C. Cassimiro, A. M. Santana, P. S. Neto, and R. L. Rabelo, "Investigating the effects of class imbalance in learning the claim authorization process in the brazilian health care market," in 2017 International Joint Conference on Neural Networks (IJCNN), pp. 3265–3272, 2017.
[61] P. Pandey, A. Saroliya, and R. Kumar, "Analyses and detection of health insurance fraud using data mining and predictive modeling techniques," in Soft Computing: Theories and Applications (M. Pant, K. Ray, T. K. Sharma, S. Rawat, and A. Bandyopadhyay, eds.), (Singapore), pp. 41–49, Springer Singapore, 2018.
[62] R. A. Sowah, M. Kuuboore, A. Ofoli, S. Kwofie, L. Asiedu, K. M. Koumadi, and K. O. Apeadu, "Decision support system (dss) for fraud detection in health insurance claims using genetic support vector machines (gsvms)," Journal of Engineering, vol. 2019, 2019.
[63] S. S.K. and V. Ilango, "A time-efficient model for detecting fraudulent health insurance claims using artificial neural networks," in 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1–6, 2020.
[64] W. Yang, W. Hu, Y. Liu, Y. Huang, X. Liu, and S. Zhang, "Research on bootstrapping algorithm for health insurance data fraud detection based on decision tree," in 2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), pp. 57–62, 2021.
[65] R. A. Bauder, T. M. Khoshgoftaar, A. Richter, and M. Herland, "Predicting medical provider specialties to detect anomalous insurance claims," in 2016 IEEE 28th international conference on tools with artificial intelligence (ICTAI), pp. 784–790, IEEE, 2016.
[66] A. Verma, A. Taneja, and A. Arora, "Fraud detection and frequent pattern matching in insurance claims using data mining techniques," in 2017 Tenth International Conference on Contemporary Computing (IC3), pp. 1–7, 2017.
[67] M. S. Anbarasi and S. Dhivya, "Fraud detection using outlier predictor in health insurance data," in 2017 International Conference on Information Communication and Embedded Systems (ICICES), pp. 1–6, 2017.
[68] H. Cao and R. Zhang, "Using pca to improve the detection of medical insurance fraud in sofm neural networks," pp. 117–122, 01 2019.
[69] S. Kareem, R. Binti Ahmad, and A. B. Sarlan, "Framework for the identification of fraudulent health insurance claims using association rule mining," in 2017 IEEE Conference on Big Data and Analytics (ICBDA), pp. 99–104, 2017.
[70] X. Jiang, S. Pan, G. Long, F. Xiong, J. Jiang, and C. Zhang, "Cost-sensitive parallel learning framework for insurance intelligence operation," IEEE Transactions on Industrial Electronics, vol. 66, no. 12, pp. 9713–9723, 2019.
[71] S. Zhou and R. Zhang, "A novel method for mining abnormal expenses in social medical insurance," in 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–5, 2020.
[72] S.-L. Wang, H.-T. Pai, M.-F. Wu, F. Wu, and C.-L. Li, "The evaluation of trustworthiness to identify health insurance fraud in dentistry," Artificial intelligence in medicine, vol. 75, pp. 40–50, 2017.
[73] V. F. de Santana, A. P. Appel, L. G. Moyano, M. Ito, and C. S. Pinhanez, "Revealing physicians referrals from health insurance claims data," Big data research, vol. 13, pp. 3–10, 2018.
[74] C. Sun, Z. Yan, Q. Li, Y. Zheng, X. Lu, and L. Cui, "Abnormal group-based joint medical fraud detection," IEEE Access, vol. 7, pp. 13589–13596, 2018.
[75] W. Liu, Q. Yu, Z. Li, Z. Li, Y. Su, and J. Zhou, "A blockchain-based system for anti-fraud of healthcare insurance," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), pp. 1264–1268, 2019.
[76] J. Gera, A. R. Palakayala, V. K. K. Rejeti, and T. Anusha, "Blockchain technology for fraudulent practices in insurance claim process," in 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 1068–1075, 2020.
[77] G. Saldamli, V. Reddy, K. S. Bojja, M. K. Gururaja, Y. Doddaveerappa, and L. Tawalbeh, "Health care insurance fraud detection using blockchain," in 2020 Seventh International Conference on Software Defined Systems (SDS), pp. 145–152, IEEE, 2020.
[78] L. Ismail and S. Zeadally, "Healthcare insurance frauds: Taxonomy and blockchain-based detection framework (block-hi)," IT Professional, 07 2021.
[79] B. Alhasan, M. Qatawneh, and W. Almobaideen, "Blockchain technology for preventing counterfeit in health insurance," in 2021 International Conference on Information Technology (ICIT), pp. 935–941, 2021.
[80] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, "A secure ai-driven architecture for automated insurance systems: Fraud detection and risk measurement," IEEE Access, vol. 8, pp. 58546–58558, 2020.
[81] G. Zhang, X. Zhang, M. Bilal, W. Dou, X. Xu, and J. J. Rodrigues, "Identifying fraud in medical insurance based on blockchain and deep learning," Future Generation Computer Systems, vol. 130, pp. 140–154, 2022.
[82] J. J. Hathaliya, S. Tanwar, S. Tyagi, and N. Kumar, "Securing electronics healthcare records in healthcare 4.0: a biometric-based approach," Computers & Electrical Engineering, vol. 76, pp. 398–410, 2019.

[83] A. A. Monrat, O. Schelén, and K. Andersson, "A survey of blockchain from the perspectives of applications, challenges, and opportunities," IEEE Access, vol. 7, pp. 117134–117151, 2019.
[84] M. McCrea and M. Farrell, "A conceptual model for pricing health and life insurance using wearable technology," Risk Management and Insurance Review, vol. 21, no. 3, pp. 389–411, 2018.
[85] B. Nayak and S. S. Bhattacharyya, "The changing narrative in the health insurance industry: Wearables technology in health insurance products and services for the covid-19 world," Journal of Health Management, vol. 22, no. 4, pp. 550–558, 2020.
[86] R. Gupta, U. Thakker, S. Tanwar, M. S. Obaidat, and K.-F. Hsiao, "Bits: A blockchain-driven intelligent scheme for telesurgery system," in 2020 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–5, 2020.
[87] wearable devices, "wearable devices will they really catch on in the health insurance industry." https://fdocuments.in/document/wearable-devices-will-they-really-catch-on-in-the-health-insurance-industry.html. Accessed: 2021.
[88] T. McGhin, K.-K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," Journal of Network and Computer Applications, vol. 135, pp. 62–75, 2019.
[89] A. Tandon, A. Dhir, N. Islam, and M. Mäntymäki, "Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda," Computers in Industry, vol. 122, p. 103290, 2020.

KHYATI KAPADIA is currently pursuing M.Tech from Institute of Technology Nirma university, Ahmedabad, Gujarat, India. Her area of specialization is data science. She is a quantum computing, blockchain, and AI enthusiast.

USHA PATEL is working as an Assistant Professor in Computer Science and Engineering Department. She has a teaching experience of more than 14 years. She has received her BTech degree in Computer Engineering from U. V. Patel College of Engineering, Hemchandracharya North Gujrat University She has pursued her M.Tech in Computer Science and Engineering from Nirma University. Currently, she is pursuing her Phd from Gujarat Technological University in the area of Image Processing. Her research interests include Image Processing and Machine Learning. She has several journal and conference papers to her credit.

RAJESH GUPTA (Student Member, IEEE) is a Full-Time Ph.D. Research Scholar in the Computer science and Engineering Department at Nirma University, Ahmedabad, Gujarat, India. He received his Bachelor of Engineering in 2008 from the University of Jammu, India and Master's in Technology in 2013 from Shri Mata Vaishno Devi University, Jammu, India. He has authored/co-authored some publications (including papers in SCI Indexed Journals and IEEE ComSoc sponsored International Conferences). Some of his research findings are published in top-cited journals and conferences such as IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, IEEE NETWORK MAGAZINE, IEEE IOT MAGAZINE, COMPUTER COMMUNICATIONS, COMPUTER AND ELECTRICAL ENGINEERING, IJCS WILEY, ETT WILEY, PHYSICAL COMMUNICATION, IEEE ICC, IEEE INFOCOM, IEEE GLOBECOM, IEEE CITS, and many more. His research interest includes Device-to-Device Communication, Network Security, Blockchain Technology, 5G Communication Network, and Machine Learning. He is also a recipient of Doctoral Scholarship from the Ministry of Electronics and Information Technology, Govt. of India under the Visvesvaraya Ph.D. Scheme. He is a recipient of student Travel Grant from WICE-IEEE to attend IEEE ICC 2021 held in Canada. He has been awarded best research paper awards from IEEE ECAI 2021, IEEE ICCCA, and IEEE IWCMC 2021. His name has been included in the list of Top 2% scientists worldwide published by the Stanford university, USA. He was felicitated by Nirma University for their research achievements in 2021. He is also an active member of ST Research Laboratory (www.sudeeptanwar.in). He is a student member of IEEE since 2018.

MOHAMMAD DAHMAN ALSHEHRI is an Assistant Professor at Computer Science Department, Taif University, Saudi Arabia and Visiting Professor at School of Computer Science at the University of Technology Sydney (UTS), Australia. He received his PhD in Artificial Intelligence of Cybersecurity for Internet of Things (IoT) from the University of Technology Sydney, Australia. He developed 6 smart novel algorithms for IoT to reinforcement Cybersecurity with AI that be able to detect the various behaviours of cyber-attacks and provide full secure and protection platform for the IoT from the most harm cyber-attacks. Furthermore, he published several publications in high ranked international journals, top-tier conferences and chapter of books, also he received number of international and national awards and prizes. His main current research interest lies in the areas of Cybersecurity, Artificial Intelligence, Internet of Things (IoT), Trust and Reputation.

SUDEEP TANWAR (M'15, SM'21)is currently working as a Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, India. He is a Visiting Professor at Jan Wyzykowski University, Polkowice, Poland; and the University of Pitesti, Pitesti, Romania. He has authored two books, edited 13 books, and more than 270 technical papers, including top journals and top conferences, such as IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE WIRELESS COMMUNICATIONS, IEEE Network, ICC, GLOBECOM, and INFOCOM. He initiated the research field of blockchain technology adoption in various verticals, in 2017. His H-index is 50. He actively serves his research communities in various roles. His research interests include blockchain technology, wireless sensor networks, fog computing, smart grid, and the IoT. He is a member of the Technical Committee on Tactile Internet of the IEEE Communication Society. He is a Senior Member of CSI, IAENG, ISTE, and CSTA. He has been awarded the Best Research Paper Awards from IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has served many international conferences as a member of the organizing committee, such as the Publication Chair for FTNCT-2020, ICCIC 2020, and WiMob2019; a member of the Advisory Board for ICACCT-2021 and ICACI 2020; the Workshop Co-Chair for CIS 2021; and the General Chair for IC4S 2019 and 2020 and ICCSDF 2020. He is serving on the editorial boards for Frontiers of Blockchain, Cyber Security and Applications, Computer Communications, the International Journal of Communication Systems, and Security and Privacy

GULSHAN SHARMA is presently working as Senior Lecturer in Department of Electrical Power Engineering, Durban University of Technology, South Africa. He is acting as DRC Chair and FRC Representative of Department of Electrical Power Engineering, Durban University of Technology. He has received Y Rated Researcher award from National Research Foundation (NRF) of South Africa. He is acting as Associate Editor of International Transactions on Electrical Energy Systems, Wiley and Regional Editor of Recent Advances in Electrical & Electronics Engineering, Bentham Science. He has more than 13 years of teaching and research experience. He has the qualifications of B. Tech, M. Tech and Ph.D. from institutes of national importance. He was a Post-Doctoral research fellow at Faculty of EBIT, University of Pretoria, South Africa from 2015 to 2016. He has published several research papers in international journals and conferences of high repute and has been continuously engaged in guiding research activities at graduate/post-graduate and Ph.D. levels. His area of interest includes power system operation and control, renewable power generation, FACTS, smart grid, hybrid electric vehicles and application of AI techniques to power systems.

PITSHOU N. BOKORO received the M.Phil. degree in Electrical Engineering from the University of Johannesburg, Johannesburg, South Africa in 2011, and the Ph.D degree in Electrical Engineering from the University of the Witwatersrand, in 2016. He is currently an Associate Professor with the University of Johannesburg, Johannesburg, South Africa. His research interests include modelling and reliability prediction of insulating materials and dielectrics, power quality, and renewable energies. He is a senior member of the South African Institute of Electrical Engineers.

RAVI SHARMA is working as a Professor in the Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun, India. Dr Sharma is passionate in the field of Business analytics and worked in various MNCs as a leader of various software development groups. Dr. Sharma has contributed various articles in the area of business analytics, prototype building for startup, and artificial intelligence. Dr. Sharma is leading academic institutions as a consultant to uplift research activities in inter-disciplinary domains.