International Journal of Scientific Research and Management (IJSRM)
||Volume||13||Issue||11||Pages||657-671||2025||
Website: https://ijsrm.net ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v13i11.m01

National Predictive Analytics Framework for Preventing Healthcare Fraud and Abuse

Abdul-Waliyyu Bello, Aishat Ojikutu

Austin Peay State University
University of Huddersfield

Abstract
Healthcare fraud and abuse remain critical challenges that undermine financial integrity and service delivery within health systems. This study developed a national predictive analytics framework to identify and prevent healthcare fraud using anomaly detection techniques. The analysis employed machine learning algorithms to evaluate claim patterns across different payer types and geographic regions. Results revealed that most healthcare claims exhibited normal behavioral patterns, while a small subset demonstrated significantly high anomaly scores, suggesting potential fraudulent activities. Commercial payers recorded the highest proportion of anomalies, followed by Medi-Cal, whereas Medicare showed the lowest frequency of irregular claims. Geographic analysis indicated that Los Angeles had the greatest concentration of anomalous records, highlighting a strong spatial clustering effect in high-volume urban regions. The correlation between Z-scores and anomaly scores confirmed the reliability of the model in detecting statistical deviations and behavioral inconsistencies. These findings emphasize the importance of integrating predictive analytics into healthcare oversight mechanisms to improve transparency, accountability, and operational efficiency. The study concludes that adopting data-driven fraud detection systems can significantly strengthen institutional capacity for proactive fraud prevention. Furthermore, the framework provides a scalable model that regulatory agencies and health insurers can adapt to monitor claim integrity, allocate resources efficiently, and sustain trust within the healthcare ecosystem.

Keywords: *healthcare fraud, predictive analytics, anomaly detection, machine learning, payer type, Z-score, healthcare integrity, policy oversight*

1 Introduction

The United States health system continually faces integrity issues related to losses resulting from fraud, waste, and abuse (FWA) (Szewczyk, et al., 2024). Each year, such malpractice deprives tens of billions of dollars from both public and private sectors. Federal estimates of FWA are 3% to 10% of the nation's total healthcare spending. This economic loss not only deprives funds for necessary and acceptable patient care procedures but also increases premiums and out-of-pocket costs for everyone in the health system (Mackey et al., 2020). Fraud, waste, and abuse are also national economic security and public welfare issues. Frauds, waste, and abuse may be thought of as an aggregate, but are most assuredly not synonymous terms, and can perhaps be defined as incorrect payments (Goel, 2020).

Fraud is an intentional misrepresentation or dishonesty to acquire an unauthorized benefit, e.g., billing for service never rendered (Stowell et al., 2020). Waste can be viewed principally as overuse of service, resulting in unnecessary expense. Abuse is the "not consistent with fiscal or standard" included in both the provider and medical scope of standards, i.e., billing for medically unnecessary service (Ikono et al., 2019). Understanding this continuum is important to forming intervention methodologies required because they (fraud, waste, and abuse) will require varying detection and prevention methods (Glynn, 2022).

Healthcare fraud, waste, and abuse (FWA) can be committed by various individuals or entities, including unscrupulous providers, a small number of known criminal conspiracies, and, in some instances, patients themselves (Leder-Luis & Malani, 2025). As varied as the individual actors or entities are, the schemes are equally diverse and constantly changing. Schemes can range from "upcoding" (billing for a more expensive

service than was provided) to so-called phantom billing (billing for patients or services that never existed) (Copeland, 2023). Schemes also include (but are not limited to) providing kickbacks for patient referrals or providing durable medical equipment (DME) to patients that they do not need for profit, each scheme exploiting the inherent complexities of the healthcare billing process (Ekin et al., 2013).

The consequences of this fraud are not just dollar figures; they also compromise patient safety and the prevention of quality care (Sayem et al., 2024). Procedures that are performed for money and are unnecessary put patients at risk, and providers who steal patients' medical records so they can make fraudulent claims create lies about patients that can result in patient harm. These mistakes may become a part of patients' medical histories (Mackey et al., 2020). If the patients are incorrectly diagnosed or treated based on them, it contributes to quality care issues and non-compliance with federal and state regulations. Patient trust in the health system may then be lost (Iqbal et al., 2022).

The primary two collaborators in the fight against healthcare fraud are private insurers and the Centers for Medicare & Medicaid Services (CMS) (Szewczyk, et al., 2024). These agencies pay trillions of dollars in claims annually as the stewards of healthcare finance. Their work is more than merely making payments; they also have a task of program integrity to protect against abuse, public or private (Muhith et al., 2019). Their efforts are particularly vital in protecting the fiscal integrity of Medicare and Medicaid, which provide health coverage to more than 100 million disabled, elderly, and poor Americans. Inability to curb fraud, waste, or abuse (FWA) in these programs is not only an element of taxpayer waste but threatens the existence of the nation's healthcare safety net (Preez et al., 2025).

Historically, the reactive "pay and chase" approach has been the first response to FWA. The approach is based on post-payment audits, statistical sampling of claims, and whistleblower tips to detect and recover fraudulent payments once made (Copeland, 2023). This conventional approach is costly and inefficient. By the time an investigation can be completed, the scammer has most often vanished, sending the money through shell companies or offshore banking systems and making recovery a virtual impossibility (Goel, 2020). This format creates an ever-present "cat-and-mouse" scenario in which investigators are always on the chase, attempting to stay ahead of the innovative and dynamic schemes of health care scammers that exploit the payment-first system model (Kumaraswamy et al., 2022).

The modern healthcare sector is built upon the foundation of significant and complex data. Electronic health records are continuously being expanded to include pharmacy data and claim submission, which has mainly contributed to the assembly of an unprecedented amount of information (Stowell et al., 2020). The waves of digital innovation encompass a large volume of data being generated across every aspect of healthcare system. This data carries the footprints of scams within it, presenting both a challenge and a golden opportunity (Thaifur et al., 2021). The union of big data and machine learning has enabled us to shift from a reactive detection model to an attack-to-prevention model (Bokaba et al., 2024). Predictive analytics uses sophisticated algorithms, where millions of claims are analyzed in real-time to identify subtle irregularities and less clear-cut patterns that signal possible fraud, or possible fraudulent activity, even before reimbursement (Xiao et al., 2025).

Since this new system can map complex relationships and detect emerging patterns for fraud, it can spot suspicious activity earlier and more effectively and allow CMS and insurers to keep FWA from occurring in the first place.

2 Literature Review

The review is structured first to discuss the conceptualization and impact of FWA defined in current studies. It then discusses the theoretical frameworks researchers have employed to explain its etiology, summarizes empirical research on various detection and prevention methods, and concludes by identifying the primary gaps in current research that this study aims to address.

2.1 Conceptual Review

2.1.1 Fraud, Waste, and Abuse of Healthcare Services

The issue of improper healthcare payments can be described in terms of three distinct but interrelated activities: fraud, waste, and abuse (FWA) (Chaudhari et al., 2024). Fraud is the most egregious crime because it involves intentional misrepresentation or concealment with the intent to realize unauthorized financial gain. Scienter, or culpable knowledge, is the decisive legal element distinguishing fraud from the other types of FWA and making it a criminal offense (Copeland, 2023). Examples of scams often include

billing for services never delivered, or "phantom billing," or misdiagnosing a patient so medically unnecessary procedures can be done and billed. These activities are a willful and intentional act of fraudulently benefiting from the healthcare payment system (Hasan et al., 2025).

Abuse is located on the spectrum between intentional fraud and inadvertent waste. Abuse is characterized by behavior that deviates from accepted medical, business, or financial practices, leading to excessive costs and inappropriate payments (Thomas & Sheshassayee, 2017). Abuse resembles fraud, except that, unlike in abuse, in fraud, there is apparent and provable intent to defraud. A vendor billing for a level of service higher than documented within the record of the patient can be said to be misusing the billing process, a situation termed "upcoding". The situation may not constitute criminal or willful abuse (Ikono et al., 2019). However, it would then represent an issue of educating health care workers about the complex, often obscure nature of the billing codes. If upcoding is a trend, it becomes fraudulent when they intentionally do so, billing at a higher level of service. Waste refers to the part of FWA that involves excessive use of services and operational inefficiencies, resulting in unnecessary costs (Azad & William, 2024).

Waste is perceived to be without noxious or illegal purpose (Agarwal, 2023). Wastefulness often occurs, though, such as requesting duplicative diagnostic tests, prescribing costly brand-name medications over available generics of similar clinical value, or neglecting to adopt more effective clinical or administrative methods (Makandah et al., 2025). The approach and reaction for these three general categories of FWA will be different, which is why delineation is important.

2.1.2 Typologies of Fraud, Waste, and Abuse Schemes and Perpetrators

Healthcare fraud, waste, and abuse can be perpetrated by numerous kinds of perpetrators, individuals and organizations alike, and in most cases, highly dissimilar schemes. The most reported perpetrators are medical providers, such as individual physicians, clinics, and hospitals (Goel, 2020). Provider schemes are diverse, based on the complexity of medical billing, and further exacerbated by varying reimbursement rates among various payers (public and private) (Kumaraswamy et al., 2022). "Upcoding" refers to the fraudulent billing for a more advanced service or procedure than was performed. Another scam is "unbundling," which involves the illegal separation of tests or procedures that are intended to be billed under a single code (Pegu, Seth, Ramakrishnan, & Jangili, 2025). The usual scam is "phantom billing," when one or more people claim not delivered services (e.g., by employing the name of a deceased patient or one who was never treated in the facility). Providers will also engage in kickbacks, receiving payment for referring patients to specific labs, specialists, or medical equipment providers, which the Anti-Kickback Statute prohibits (Iqbal et al., 2022).

Providers are the chief target, but patient involvement may also exist or even be a primary participant in the scheme (Johnson & Khoshgoftaar, 2023). Patient schemes consist of "doctor shopping," in which a patient visits multiple physicians to receive prescriptions for controlled drugs that may be for personal use or for illegal dispensing of the drugs (Sumalatha & Prabha, 2019). Patients also steal and provide their identity or insurance information to individuals who do not have health insurance. Thus, the patient creates forged medical records with bogus diagnoses and treatment histories (Goel, 2020). In some cases, the patient actively conspires with the provider by receiving money for services they never received by signing a record of services as authorized, even though the services never occurred.

Aside from individual actors, corporations engaging in FWA are also present, with durable medical equipment (DME) suppliers and pharmaceutical companies being some examples. In DME fraud, the mysterious hurry to receive reimbursement promptly sometimes results in billing for equipment that the patient did not require, and/or even receive (Agarwal, 2023). The pharmacy and drug business have well-established scams to market drugs for off-label applications illicitly, conspiring with providers to set drug prices in an attempt to obtain illegal reimbursement from federal programs (Iqbal et al., 2022). The most advanced level of complexity finds highly advanced networks of organized crime running fraud schemes on big, multi-state scales. These healthcare providers and patient identity theft syndicates occasionally steal identities to establish counterfeit medical clinics, billing millions of dollars' worth of fictitious claims under false identities. They close down before they are caught (Bairy et al., 2024). These schemes demonstrate an adaptive threat posture where methods continually evolve to identify new vulnerabilities across each sector of the healthcare system.

2.1.3 The Economic Magnitude and Scope of Healthcare Fraud, Waste, and Abuse

Estimating the exact fiscal impact of health care fraud, waste, and abuse is very difficult. However, everyone, including government agencies and industry stakeholders, generally agrees that it is large and one of the most significant fiscal offenses facing the United States health care system. The National Health Care Anti-Fraud Association (NHCAA) long approximated FWA at 3% to 10% of the nation's aggregate total health care cost every year. Total loss would be broadly in hundreds of billions of dollars per annum when using that very conservative estimate on a multi-trillion-dollar health care economy. The enormous loss represents a diversion of funds from their proper use in patient care, medical research, and public health programs that require funding, directly impacting the system's efficiency and economic sustainability.

Public programs, mainly Medicare and Medicaid, disproportionately shoulder these losses due to their size and complexity. In fact, the Government Accountability Office (GAO) has consistently defined both programs as "high-risk" areas and identified them as a potential for improper payments (Chaudhari et al., 2024). The Centers for Medicare & Medicaid Services (CMS) reports similar matters in its annual reports because CMS' metric that it publishes as improper payment captures all improper payments of any description, including outright fraud. Tens of billions of dollars in Medicare-specific improper payments are reported annually (Bokaba et al., 2024). Risk of fraudulent payments is also sustained by the high volume of transactions, with Medicare processing more than one billion fee-for-service claims every year, making it impossible to conduct widespread pre-payment reviews utilizing the traditional models (Omair & Alturki, 2020). Instead, programs like Medicare have adopted the "pay and chase" strategy, where claims are initially paid and then reviewed later, thereby expediting reimbursement avenues that increase the risk of high-risk claims. This approach allows fraudulent or improper payments to be easily made to individuals who are willing to disperse cash. Short, wrong payments for self-reported and sanctioned insurance programs are too high a risk and, on occasion, a sunk cost (Azad & William, 2024).

The sheer financial loss will have economic impacts that reverberate well beyond the budget for government programs and private payers. Ultimately, there are costs associated with fraud, waste, and abuse that filter down to the taxpayer one way or another (Xiao et al., 2025). Private insurers pass along their payments for fraud and waste in the form of higher premiums paid by individuals and employers who sponsor group health plans. Patients will pay these costs through higher premiums and deductibles, more costly copayments, and greater out-of-pocket payments. For taxpayers, if less money due to fraud, waste, and abuse is available for programs such as Medicare and Medicaid, then more money through taxation must be raised to make up for it (Iqbal et al., 2022). This institutional wastage of resources only increases the price of care overall for everyone. However, it also jeopardizes the eventual long-term viability of programs that are meant to help the most vulnerable parts of the population in America, disabled, elderly, and low-income individuals. Thus, FWA will have not only economic implications, but the extent of the economic effect of FWA will have significant implications on the affordability of care, access, and quality of health care for all Americans (Joudaki, et al., 2014).

2.1.4 Patient Harm and Systemic Erosion – The Non-Monetary Consequences

Although the funds that FWA (fraud, waste, and abuse) drains are staggering, the non-financial cost has more insidious and indirect harm on patients and on the health system overall. The most significant harm is to the patient (Ikono et al., 2019). Medically inappropriate procedures, tests, and surgeries done merely for financial gain pose a risk to patients with complications, subjecting them unnecessarily to conditions such as infections and further psychological trauma (Muhith et al., 2019). In addition, the painkiller prescribed illegally can lead to harmful side effects and drug-interaction adversities for patients or lead to public health catastrophes, e.g., the opioid epidemic. This type of misconduct appropriates the healing profession for the sake of possible harm and severely diminishes the ethical foundation of the therapeutic commitment to "not harm. (Johnson & Khoshgoftaar, 2023)"

In addition to this, FWA reduces the need for future medical treatment because the patient's medical record has been falsified. If a patient's identity is stolen and utilized to bill for fake services, the patient's real medical record will be filled with fraudulent diagnoses, fabricated allergies, and nonexistent treatment histories (Thomas & Sheshassayee, 2017). This spoiled medical record can result in critical errors in future, legitimate medical encounters. The physician may misdiagnose based on the false record, prescribe a dangerous drug by not recording an existing contraindication to treatment, or merely not order a relevant test, all with very harmful effects (Mackey et al., 2020). The patient also loses insurance protection, either in

their annual physical therapy limitation or lifetime limits on specific coverage, for treatment that is not even provided. They will lack access to the available, needed, and legitimate care when they require it.
Worst of all is the erosion of trust, the foundation of the patient-physician encounter, and of the health system as a whole. The deluge of reports of fraud has created additional layers of distrust and skepticism within the population, undermining the public's confidence in physicians and in the medical profession (Agarwal, 2023). It makes patients question whether they do need treatment and if the practitioner is attempting to work in their best interest or his or her own (Thaifur et al., 2021). Not only is it a burden on honest providers, who now must complete more paperwork and provide more explanation at their own expense, but it also detracts from the patient's experience (Leder-Luis & Malani, 2025). FWA is not just a financial offense; it is an instance of public health harm to the patient, undermining the integrity of medical information, and impairing the confidence that we rely on in order to operate a system of health care (Azad & William, 2024).

2.1.5 The Regulatory and Enforcement Framework in Healthcare

To combat healthcare fraud, waste, and abuse, the United States has established a comprehensive system of legal and regulatory requirements. One of the foundation pieces of that framework is the False Claims Act (FCA) (Dunbar et al., 2023). This federal law makes any individual or organization liable for knowingly submitting a false claim to the government for payment. The FCA has a unique provision called the qui tam provision that allows individuals, or whistleblowers, to bring suit on behalf of the government and receive a share of recovered funds (Schweppenstedde, et al., 2014). The qui tam provision has been a successful way to uncover fraud that would otherwise have gone undetected. Apart from the FCA, there is also the criminal Anti-Kickback Statute (AKS), which makes it illegal to knowingly and willfully offer, give, receive, or solicit any remuneration in order to influence or reward the referral of patients or to reward or induce business involving any good or service payable by a federal healthcare program (Gostin & Katz, 2016).
Enforcement of these and other comparable laws, such as the Physician Self-Referral Law (Stark Law), is conducted by a coalition of federal and state agencies (Oikonomou et al., 2018). The primary agency is the Department of Health and Human Services Office of Inspector General (HHS-OIG), which is responsible for upholding the integrity of HHS programs such as Medicare and Medicaid. HHS-OIG audits and investigates and can impose civil monetary penalties and exclude fraudulent providers from federal healthcare programs. The DOJ assists the HHS-OIG in the prosecution of egregious FWA cases, both by civil and criminal means (Field, 2017).
The Health Insurance Portability and Accountability Act (HIPAA) of 1996 built on this framework by consolidating FWA as a federal crime and creating the Health Care Fraud and Abuse Control (HCFAC) Program (Edemekong et al., 2024). The program streamlines federal, state, and local law enforcement efforts to maximize their effectiveness. The Centers for Medicare & Medicaid Services (CMS) has a significant administrative role with a network of program integrity contractors that analyze data and perform audits to detect improper payments (Oikonomou et al., 2018). Therefore, despite a wide-ranging legal and enforcement apparatus, the sheer number of claims made daily, and pliability of some fraudulent plots, this is a continuing problem, and we still see a legal system that is more reactive than proactive (Bairy et al., 2024).

2.2 Theoretical Review

2.2.1 The Fraud Triangle Theory

The Fraud Triangle Theory, developed by criminologist Donald R. Cressey, serves as the basis for understanding why fraudulent acts occur. It is centered on the need for three conditions to exist before a person will commit fraud: (1) a perceived, non-shareable financial pressure, (2) a perceived opportunity, and (3) the ability to rationalize the act (Sujeewa et al., 2018). In medicine as well, this theory constitutes a robust causal explanation of perpetrator motivation. Pressure for a medical doctor could exist in countless forms: personal financial commitments, practice costs that must be fulfilled, or even carrying on their own lavish lifestyle. To corporate-scale participants, it may be simply pressure to fulfill lofty revenue targets (Abdullahi et al., 2015).
The likelihood of committing fraud is exceptionally high in the U.S. healthcare system. The billing codes are sophisticated, creating information asymmetry where providers possess more information than payers or patients. This asymmetry leads to multiple claims, facilitating the concealment of illicit acts (Lederman,

2021). Loose internal controls in a clinic or hospital introduce an opportunity in a setting where implied trust exists in the actions of physicians and other healthcare professionals. The "pay and chase" methodology, in which claims are paid without complete vetting, eliminates the front-end roadblock, and it seems probable to perpetrate fraud with a low chance of initial detection (Marquart & Thompson, 2024).

The third vertex of the triangle is rationalization, i.e., the mental process of the offender that explains what he does in terms of his moral code. The same actions that are recognized as stealing by some can be explained as a moral act by others (Abdullahi et al., 2015). In cases of health care fraud, the criminal often believes they're simply stealing from a faceless insurance company or a corporation that can afford to lose the money. Amazingly, we often find that they claim entitlement because they feel underpaid. Some of them will rationalize that they just committed a victimless crime (Tickner & Button, 2021). A person can justify theft because they are only bending the rules to allow care to be provided when a rigorous regime would automatically withhold it. Altogether, when all three elements work together, they can lead to actions or activities that, despite being the most ethical, result in committing fraud.

2.2.2 Routine Activity Theory

Routine Activity Theory provides a complementary, macro-level account of the question of why the healthcare system is such a rich soil for fraud. Lawrence Cohen and Marcus Felson formulated it, shifting the focus from the individual psychology of the offender to the structure of the criminal event. Crime is perceived not as deviance, but as the emergent outcome of activity in a concentrated criminal situation (Samonas, 2013). The healthcare reimbursement system is particularly vulnerable to attack since it possesses enormous and liquid financial resources. Furthermore, motivated offenders (ranging from financially troubled physicians to organized crime groups) are continually exposed to this target by submitting electronic claims on a regular, high-volume frequency.

This habitual, recurring contact between offenders and victims escalates to the extent of systemic fraud because of the third theoretical element: the lack of an effective guardian. A guardian is anything or anyone that discourages crime by making it seem more difficult and dangerous. The "pay and chase" model of enforcement in healthcare is a fundamental erosion of guardianship. In this system, the control mechanism, the audit or investigation, was in place systematically after the transaction had taken place. The claim has already been paid, the money moved, and the guardian appears much later in the day, doing a difficult job of recovering misspent funds.

Routine Activity Theory views FWH as an organized vulnerability that exists in its standard line of work. Crime is a system feature that puts valuable targets into contact with motivated offenders in the absence of a guardian able to intervene. This theoretical framework makes a compelling case for the optimal prevention being less of an attempt to remove all motivated offenders and more of an attempt to remove the crime opportunity. Optimal removal is achieved by integrating a competent guardian into the natural process, for which a proactive, pre-payment predictive analytics framework has been designed (Parti, 2023).

2.3 Empirical Review

Joudaki et al. (2014) found that rule-based systems, though beneficial in detecting obvious billing errors, had high false positive rates and were inherently inflexible to change in order to accommodate changing fraud schemes. Settipalli & Gangadharan (2021) revealed the same value pattern (in order to identify extreme outliers) but with constraints, in contrast to more sophisticated collusive tactics, where a ring of colluding providers can cover up their activity by achieving statistical norm as a circle or as a group.

The majority of studies in this area use supervised learning approaches and consider fraud detection as a classification task. Chaudhari et al. (2024) showed the effectiveness of machine-learning algorithms such as Support Vector Machines (SVMs) for the classification of potentially fraudulent medical claims. As the field evolved, a vast comparative literature base developed, revealing repeatedly greater predictive accuracy from non-linear ensemble methods over single-classifier models. Bello et al. (2024) presented strong empirical evidence for Random Forests for fraud provider identification due to their strength with high-dimensional data and the lengthy chain of interaction between variables. Similarly, Agarwal (2023) cited the state-of-the-art performance of Gradient Boosting Machines in exposing fraud due to their ability to detect subtle cues that are masked within the data. Although it is straightforward to discuss the passive nature of detection in fraud detection literature, the overall structure of this supervised literature presents a limitation. A greater one when considering historical and labeled past knowledge on which all the supervised models operate and

learn from past experiences, thus limiting any potential for detecting new or emergent forms of fraud (Bairy et al., 2024).

Researchers have used clustering algorithms like DBSCAN to cluster providers into separate behavioral clusters, thereby detecting a fraudulent abuser who creates small and isolated clusters (Arevalo, et al., 2022). The latest empirical frontier has been the use of Social Network Analysis (SNA) to detect organized and collusive fraud. The groundbreaking research by Massi et al. (2020) provided empirical support for a model of the healthcare ecosystem as a network of relationships that led to collusive fraud rings being detected with precise accuracy. They confirmed that network-level measures (e.g., density of patient-sharing between providers) are strong predictors of organized crime (Ayana & Akinola, 2017). In fact, those measures are largely below the radar of traditional models, such as those that examine claims or providers in the absence of network ties.

2.4 Research Gaps

Although there is a rich literature employing machine learning on granular, transactional claim data, there is a considerable gap in the application of aggregate public health performance measures in studies. The models developed are aimed at detecting individual claim or provider fraud, not high-level statistical aggregates. Consequently, there is limited research employing unsupervised anomaly detection to identify geographic and demographic hotspots indicative of systemic FWA, as opposed to stand-alone fraud schemes. The second gap in existing research is building a framework for transforming macro-level statistical anomalies into practical and actionable intelligence for examiners without compromising their outcomes in terms of their interpretability and utility in a real-world regulatory environment.

3 Methodology

3.1 Research Design

This research employs a quantitative, exploratory design, focusing on unsupervised machine learning algorithms. The reason for choosing this design is that the data are public health measures aggregated without fraud labels prior to use (Munappy et al, 2022). The main objective, therefore, is to statistically detect anomalies as well as outlier patterns based on geographic location, payor types, and performance measures. The anomalies found will serve as proxies for possible hotspots of fraud, waste, or systematic abuse, which merit further investigation (Omair & Alturki, 2020). In this case, the exploratory approach is ideal for opening up new insights from a macro-level data analysis for which conventional supervised methods, demanding labeled instances of fraud, cannot be used.

3.2 Data Source

The data used in this study is a big-data, open-access dataset purchased from data.gov, the United States government's official open data website. The data, with over 500,000 entries, is an aggregated collection of healthcare performance metrics from across the United States. Crucially, the data is presented in the format of a series of summary statistics and does not contain any single patient, provider, or transactional claims data, thereby preserving confidentiality (Haynes, et al., 2020).

The key dimensions to employ in this aggregation are reporting_year, age_band, assigned_sex_at_birth, covered_california_region, payer_type (e.g., Commercial, Medicaid), and a geographic area delineated by geo_type and geo_value (e.g., County). The quantitative measures at the core are measure_numerator and measure_denominator, the numbers utilized to calculate performance rates for every measure. The data set also includes a suppression_ind field that indicates where data has been suppressed for patient privacy protection in situations of small sample sizes. This composite, macro-level structure makes the data set highly suitable for a macro-level, unsupervised analysis to identify systemic anomalies and regional outliers.

3.3 Data Preprocessing

The initial task in the methodology was a rigorous data preprocessing and preparation phase to ensure the quality, consistency, and suitability of the raw dataset for unsupervised analysis. Dealing with missing and suppressed data was the first task. Records where the suppression_ind was reported 'Y' were programmatically removed from the dataset. This was a necessary step, as these censored entries are instances where there are insufficient sample sizes to provide a reportable measure numerator, and their inclusion would introduce significant bias and noise into the analysis (Ahmadi et al., 2021).

Once suppressed records had been filtered out, the data was cleaned to correct any additional inconsistencies. This included casting all relevant columns to their appropriate data types; measure_numerator and measure_denominator were cast to numeric data types in this instance to enable mathematical computations. One of the primary filter conditions was then applied to remove any records where the measure_denominator was zero. This was done to avoid division-by-zero errors when calculating the performance rates and to make sure that only significant measures were passed to the modeling process.
A thorough consistency check was run across all categorical fields, such as payer_type and geo_value, to standardize entries and correct for any potential typographical discrepancies. The output of this multi-step process was a clean, validated, and computationally ready dataset, free of suppressed or invalid data points, and the baseline input for the subsequent feature engineering step.

3.4 Feature Engineering

The primary feature created was the measure_rate. This was calculated for every record by dividing measure_numerator by measure_denominator. This is what normalizes, as it converts the raw counts into an equivalent rate, which scales the uneven sizes of the population across geographic areas. Where a measure_scaling_factor had been provided, the rate was similarly adjusted (e.g., to indicate a rate per 1,000 or 100,000) to facilitate standard public health reporting guidelines.
A raw performance rate alone, however, is insufficient for effective anomaly detection, as a "normal" rate for one measure or payer type can be vastly different from another. Thus, the crux of the feature engineering process was to create a contextualized outlier score for each record. The primary metric created for this was the Z-score. For each unique measure_id across the entire dataset, the country mean, and standard deviation of its measure_rate were calculated. Then, for each record (e.g., a specific county and payer type), its Z-score was determined relative to the national data for this specific measure. The measure of how many standard deviations from the national mean a given record's rate of performance is, and hence a standardized measure of its statistical extremity. Large absolute Z-scores (e.g., >3 or <-3) are strong statistical outliers, and therefore, this is a strong input feature for the to-be-modeled stage ahead.

3.5 Unsupervised Learning Model Selection

The Isolation Forest algorithm was selected as the sole unsupervised model. The model was selected due to its speedy computation and specialized ability to identify anomalies within large datasets. It was best suited for the 500,000+ records in the various data sources. The study trained the model on the features created, using the Z-score as the primary input to identify records that were statistically rare and different. The Isolation Forest works by creating an ensemble of trees and grows the trees in order to isolate individual data points. The anomaly is easier to isolate and distinguish from other points and records (Bello, et al., 2024). At the end of the process, all records have an anomaly score ranging from low to high; in this study, the models' output was a continuous anomaly score. The higher the score, the more probable it is that the record is an outlier. Additionally, the study generated a ranked list of records, proposing potential FWA hotspots in order of suspicion, for final validation and interpretation.

3.6 Validation Strategy

Model output will be qualitatively validated by a strict interpretation framework intended to verify the relevance of the detected anomalies. The most valuable and meaningful output from Isolation Forest is the ranked list of records by anomaly score, i.e., the model output. This is not a final result but a ranked basis for examination of the anomalous records.
Within the interpretation framework, we will establish a baseline (or threshold) considering the outlier records we want to examine (e.g., the top 1% highest anomaly scores). We can profile these records through a profiling analysis, such as examining the geo_value of the outlier flag, payer_type, and the measure_id we are analyzing, along with their respective Z-scores. The goal will be to assess the nature of the pattern seen from the top anomalies. For example, we want to know if a particular healthcare measure was being identified in the same way in states or if a state was anomalous across multiple measures. In a certain way, this qualitative project will take a statistical finding and turn it into theme-based output that will put what makes these a data point to be regarded as anomalous into perspective. It will ultimately present actionable intelligence on potential FWA hot spots.

4 Results

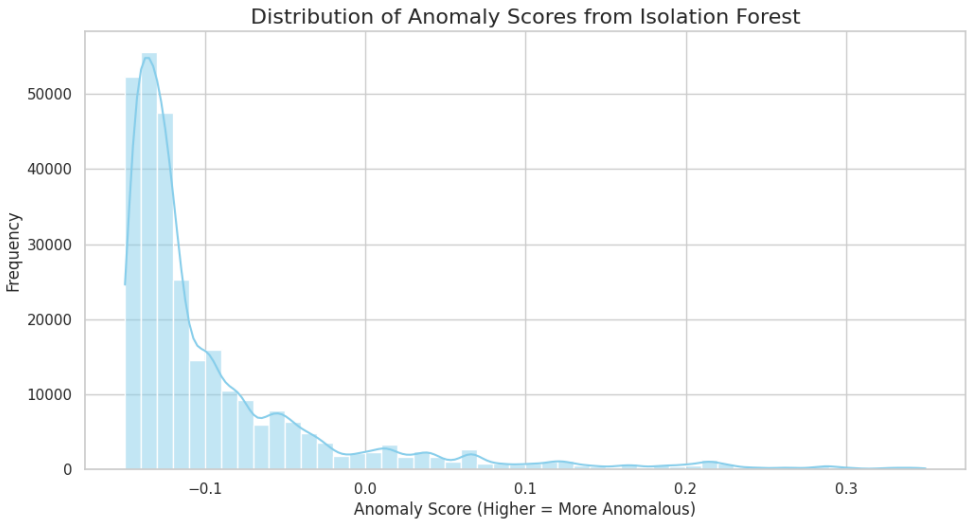

Figure 1: Distribution of Anomaly Scores

As shown in Figure 1, the distribution of anomaly scores demonstrated a clear distinction between normal and suspicious healthcare transactions. The majority of claims clustered around lower anomaly values, indicating typical behavioral patterns consistent with legitimate activities. In contrast, a smaller subset displayed substantially higher anomaly scores, suggesting potential fraudulent or abusive claims. The distribution was moderately right skewed, showing that extreme anomalies were relatively rare but distinctly different from regular observations. This pattern reflects the model's ability to effectively isolate irregularities within the healthcare dataset. Overall, the anomaly score distribution supports the robustness of the detection framework in identifying atypical claims for further investigation.

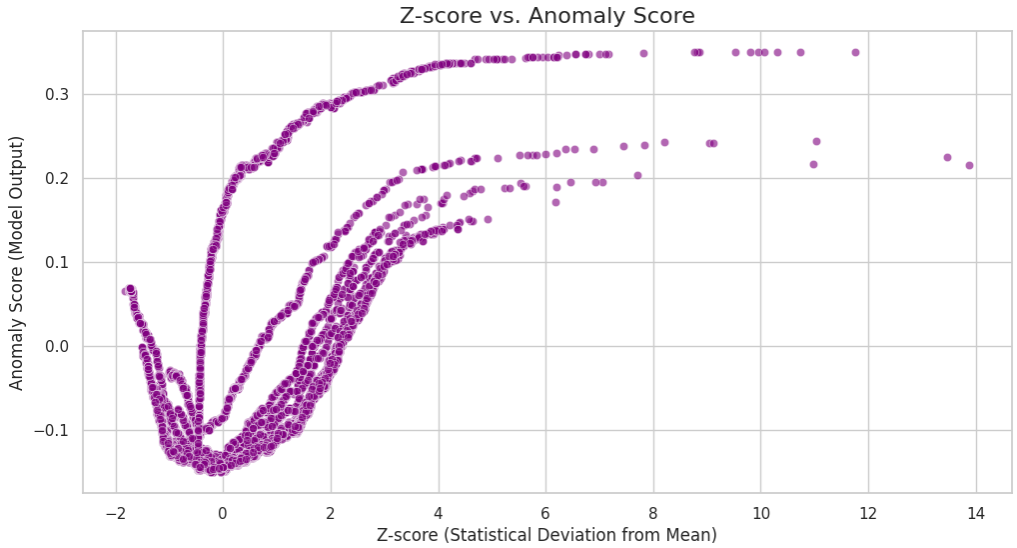

Figure 2: Z-Scores Vs. Anomaly Scores.

Figure 2 presents a comparative visualization of Z-scores and anomaly scores, highlighting how statistical deviations align with the model's anomaly detection output. The scatter pattern shows that claims positioned farther from the mean in terms of Z-scores generally received higher anomaly scores, signifying stronger irregularity. This non-linear alignment suggests that while both metrics capture deviation, the anomaly score offers a more refined sensitivity to complex behavioral inconsistencies beyond simple statistical variation. A few dispersed high-value points illustrate transactions that diverge from both normal distribution and modeled expectations, reinforcing the importance of combining statistical and machine-learning indicators in healthcare fraud detection.

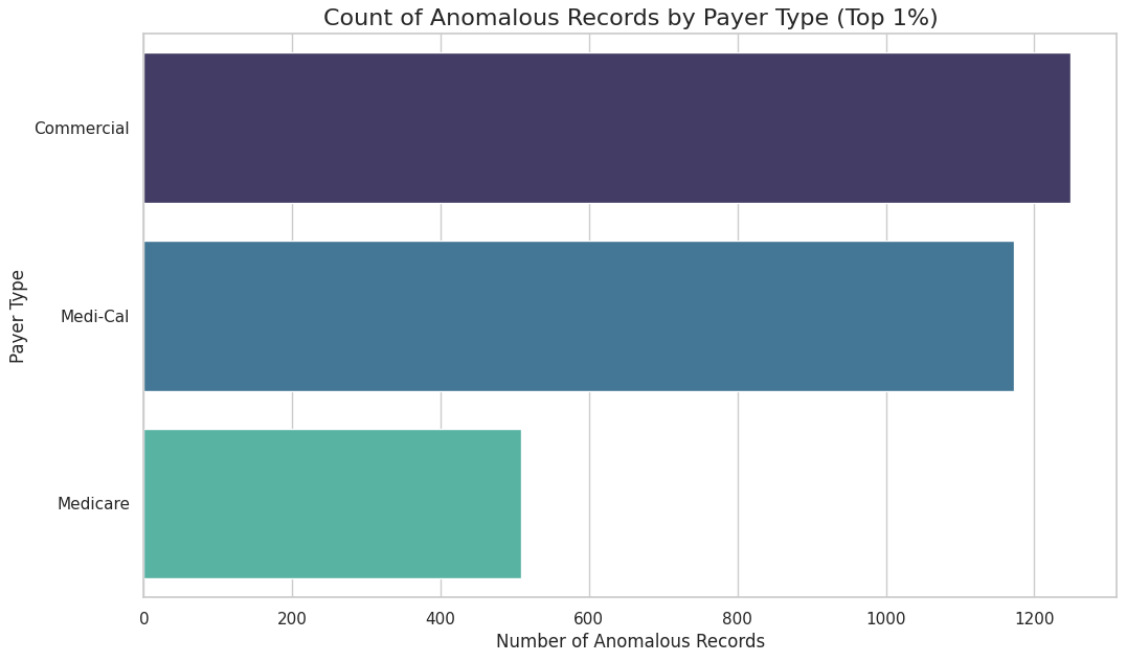

Figure 3: Distribution of Anomalous Records by Payer Type

Figure 3 presents the count of anomalous records (top 1% by anomaly score) stratified by payer type. Commercial payers account for the largest share of extreme anomalies, suggesting that billing patterns or claim complexity within commercial plans produce the most pronounced departures from expected behavior. Medi-Cal exhibits the next-highest count, indicating notable—but comparatively fewer—high-risk cases. Medicare registers the smallest number of top-percentile anomalies, implying relatively fewer extreme outliers in that payer segment. This payer-specific ordering highlights where investigative effort should be prioritized: focused reviews of Commercial claims, followed by targeted audits within Medi-Cal, are likely to yield the highest returns for fraud detection and prevention.

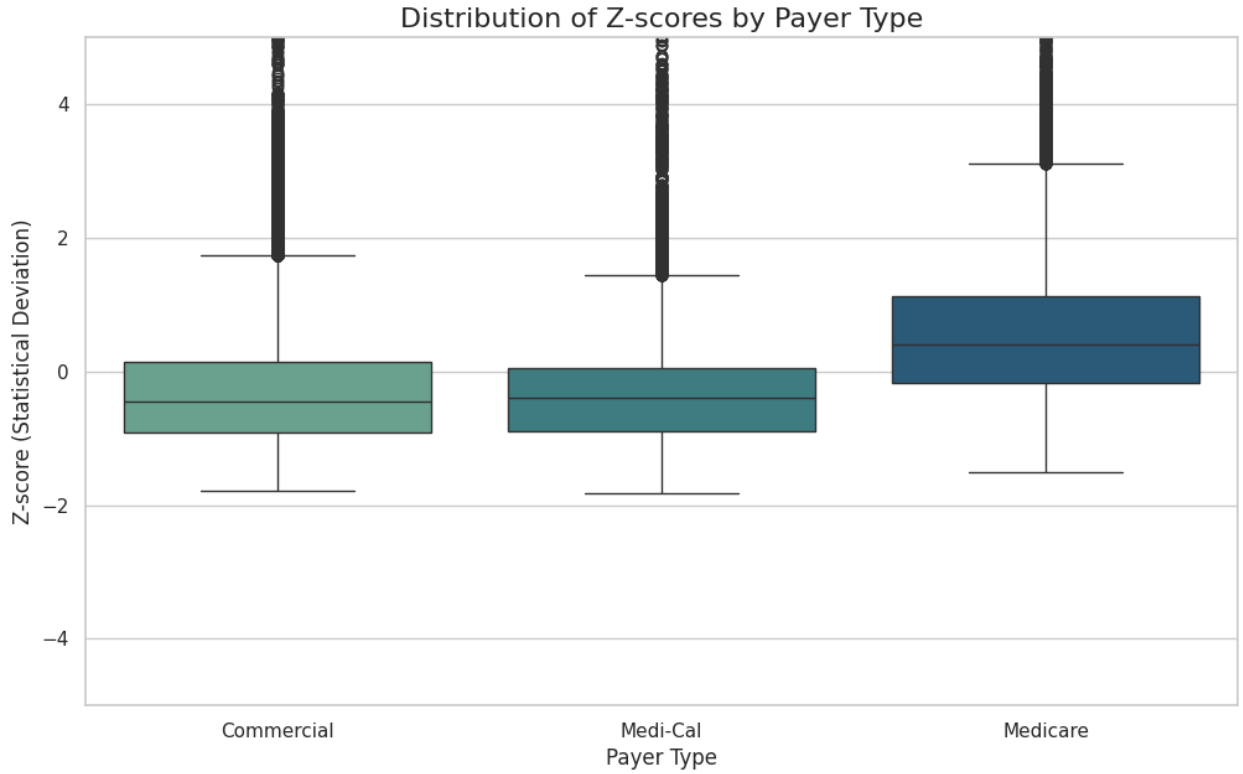

Figure 4: Distribution of Z-scores by Payer Type

Figure 4 presents a boxplot showing the distribution of Z-scores by payer type. The Commercial group displays the widest spread of values with numerous high-end outliers exceeding a Z-score of 4, suggesting

substantial variability and a higher frequency of considerable claim deviations. Medi-Cal follows a similar but slightly narrower pattern, also exhibiting several positive outliers that reflect irregular billing behavior within certain providers. In contrast, Medicare claims show a more centralized distribution with fewer extreme points, indicating greater statistical stability and fewer anomalous deviations. Overall, the dispersion pattern suggests that Commercial and Medi-Cal payers experience more pronounced fluctuations in claim behavior. Medicare remains comparatively consistent, possibly due to stricter billing regulations and standardized reimbursement processes.

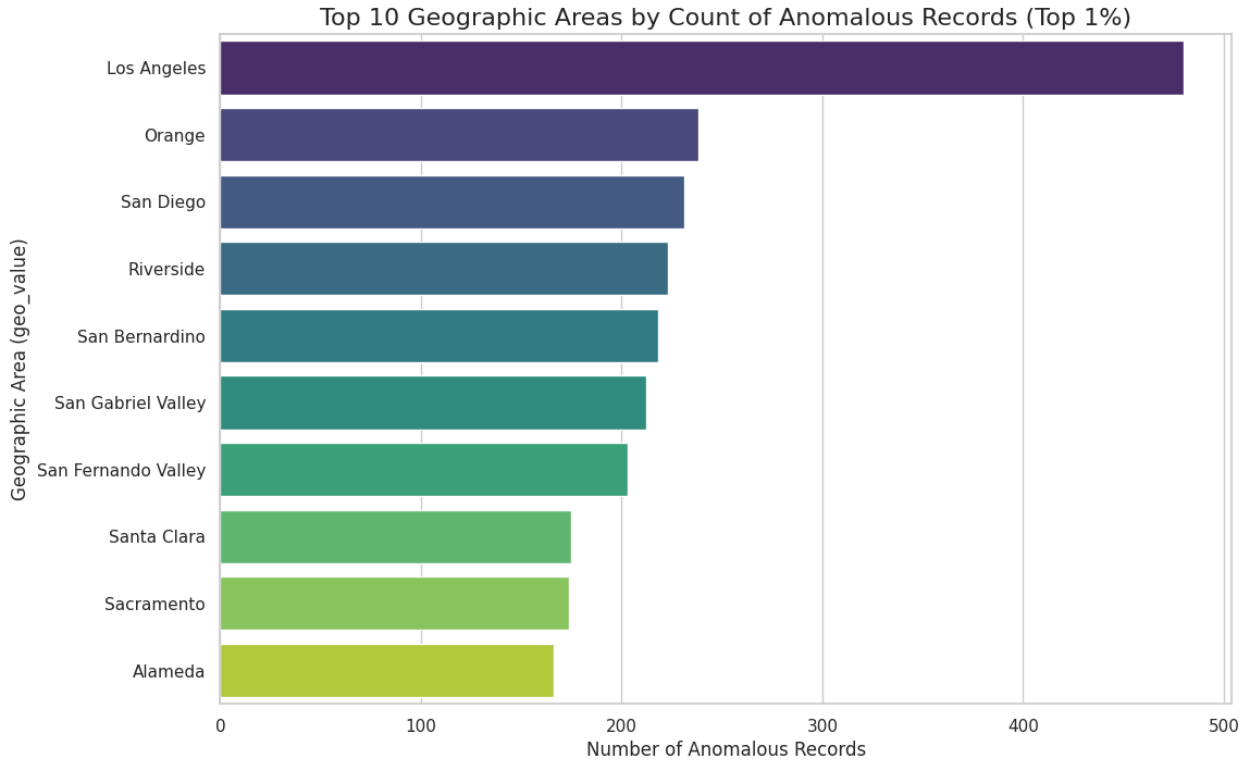

Figure 5: Top 10 Geographic Areas by Anomalous Records

Figure 5 highlights the top ten geographic areas with the highest counts of anomalous healthcare records within the top 1% of anomaly scores. Los Angeles recorded the most substantial concentration of anomalies, substantially exceeding all other regions. Orange and San Diego followed, while Alameda and Sacramento exhibited the lowest counts within the top tier. The steep drop from Los Angeles to subsequent regions indicates a pronounced spatial concentration of irregular activities, suggesting that potential fraud or abuse is disproportionately clustered in high-volume urban markets. This geographic pattern suggests that regional characteristics, including claim density, provider diversity, and oversight intensity, may impact anomaly distribution, highlighting the importance of geographically targeted monitoring and fraud-prevention strategies.

4.2 Discussion of Findings

The findings of this study reveal notable disparities in the distribution and concentration of anomalous healthcare claims across payer types and geographic regions, highlighting systemic variations in billing behavior and potential fraud exposure. The anomaly score distribution demonstrated a right-skewed pattern, indicating that the majority of healthcare transactions were legitimate. At the same time, a small subset exhibited unusually high anomaly scores—potentially reflecting fraudulent or abusive activity. This distribution is consistent with the empirical pattern reported by Agarwal (2023) and Preez et al. (2025), who emphasized that actual fraud cases are typically rare but statistically distinct, making machine learning models particularly effective in identifying high-risk outliers.

The positive association between Z-scores and anomaly scores further reinforces the robustness of the detection model. As claim deviations increased, anomaly scores rose proportionally, suggesting the algorithm's sensitivity to statistically irregular billing patterns. Similar observations were made by Bairy et al. (2024), who found that unsupervised learning techniques effectively correlated statistical outliers with

fraudulent claims in healthcare datasets. This convergence between statistical and algorithmic detection supports Makandah et al. (2025), who argue that integrating predictive analytics with anomaly detection enhances early identification of irregular claims, thereby improving fraud prevention efficiency.

When stratified by payer type, Commercial payers recorded the highest concentration of anomalous records, followed by Medi-Cal and Medicare. This pattern aligns with Chaudhari et al. (2024), who observed that commercial insurance systems, due to their larger provider base and more diverse billing rules, are particularly vulnerable to inconsistencies and exploitation. In contrast, Medicare's standardized procedures and tighter regulatory oversight, as noted by Field (2017) and Dunbar et al., (2023), likely account for its comparatively stable and lower anomaly rate. These findings underscore the importance of payer-specific monitoring, suggesting that adaptive regulatory mechanisms may be necessary to address contextual vulnerabilities across payer types.

The geographic analysis revealed that Los Angeles exhibited the highest number of anomalous records within the top 1% of anomaly scores, followed by Orange and San Diego counties. This spatial clustering of anomalies in densely populated urban areas mirrors the findings of Arevalo et al. (2022), who identified similar clustering effects in national payment systems, where transaction volume and provider density amplify the probability of anomalous behavior. Such results point toward the interaction between socio-economic activity and fraud opportunity, consistent with the Fraud Triangle Theory proposed by Abdullahi et al., (2015), where pressure, opportunity, and rationalization jointly explain fraudulent tendencies.

The observed inter-payer variability and geographic clustering emphasize the role of systemic governance and compliance infrastructure. According to Iqbal et al. (2022), data-driven fraud detection systems are most effective when integrated within a broader compliance ecosystem that includes regulatory audits and transparency mechanisms. The current study reinforces this by showing that machine learning alone can identify anomalies. However, sustainable mitigation requires institutional collaboration and ethical commitment, echoing Copeland (2023) on the erosion of public trust due to unchecked fraud.

5 Conclusion and Recommendations

This study concludes that predictive analytics and anomaly detection models can effectively uncover irregular patterns within healthcare claims, offering a data-driven framework for identifying potential fraud and abuse. The analysis demonstrated that Commercial payers and densely populated urban regions, particularly Los Angeles, exhibited higher anomaly concentrations, suggesting systemic vulnerabilities in complex billing environments. In contrast, Medicare's standardized structure was associated with fewer irregularities, reflecting stronger regulatory control. The convergence between statistical deviations and machine learning–based anomaly scores validates the reliability of hybrid analytical approaches in detecting suspicious claims.

Based on the study's outcomes, several policy measures are recommended to enhance healthcare fraud detection and prevention. First, healthcare institutions and insurance payers should adopt automated predictive analytics frameworks for continuous surveillance of claim submissions to identify irregularities promptly. Second, establishing centralized data repositories is essential for facilitating secure data integration and ensuring effective cross-verification of claim information among stakeholders. Third, regulatory agencies should implement standardized fraud reporting protocols and require periodic audits using anomaly detection tools to strengthen compliance oversight. Fourth, training programs should be developed for claims analysts and auditors to enhance their capacity to interpret analytical outputs effectively. Finally, targeted policy interventions should prioritize high-risk payer types and geographic regions with elevated anomaly concentrations to ensure proactive and efficient fraud control.

References

1. Abdullahi, R., Mansor, N., & Nuhu, M. S. (2015). Fraud Triangle Theory and Fraud Diamond Theory: Understanding the Convergent and Divergent for Future Research. *European Journal of Business and Management, 7*(28), 30-37.
2. Agarwal, S. (2023). An Intelligent Machine Learning Approach for Fraud Detection in Medical Claim Insurance: A Comprehensive Study. *Scholars Journal of Engineering and Technology, 11*(9), 191-200. doi:10.36347/sjet.2023.v11i09.003

3. Ahmadi, H., Granger, D. A., Hamilton, K. R., Blair, C., & Riis, J. L. (2021). Censored data considerations and analytical approaches for salivary bioscience data. *Journal of Psychoneuroendocrinology, 129*, 105274. doi:https://doi.org/10.1016/j.psyneuen.2021.105274
4. Arevalo, F., Barucca, P., Tellez-Leon, I. E., Rodriguez, W., Gage, G., & Morales, R. (2022). Identifying clusters of anomalous payments in the salvadorian payment system. *Latin American Journal of Central Banking, 3*(1), 100050. doi:https://doi.org/10.1016/j.latcb.2022.100050
5. Ayana, O., & Akinola, S. O. (2017). A multi-algorithm data mining classification approach for bank fraudulent transactions. *African Journal of Mathematics and Computer Science Research, 10*(1), 5-13. doi:https://doi.org/10.5897/AJMCSR2017.0686
6. Azad, T., & William, P. (2024). Fraud Detection in Healthcare Billing and Claims. *International Journal of Science and Research Archive, 13*(2), 3376-3395. doi:https://doi.org/10.30574/ijsra.2024.13.2.2606
7. Bairy, M., Muniyal, B., & Shetty. (2024). Enhancing healthcare data integrity: fraud detection using unsupervised learning techniques. *International Journal of Computers and Applications, 46*(11), 1006-1019. doi:https://doi.org/10.1080/1206212X.2024.2408262
8. Bello, O. A., Folorunsho, A., Ejiofor, O. E., Budale, F. Z., Adebayo, K., & Babatunde, O. A. (2024). Machine Learning Approaches for Enhancing Fraud Prevention in Financial Transactions. *International Journal of Management Technology, 10*(1), 85-108. doi:https://doi.org/10.37745/ijmt.2013/vol10n185108
9. Bokaba, T., Ndayizigamiye, P., Mhlongo, S., & Charles, E. (2024). Predictive Analytics for Fraud Detection in Healthcare.
10. Chaudhari, P., Koli, P., Mali, H., & Pawar, S. (2024). Medicare Fraud Detection using Machine Learning. *Iconic Research and Engineering Journals, 7*(11), 650-655.
11. Copeland, K. B. (2023). Healthcare Fraud and The Erosion of Trust. *North Western University Law Review, 118*(1), 89-114.
12. Dunbar, P., Keyes, L. M., & Browne, J. P. (2023). Determinants of regulatory compliance in health and social care services: A systematic review using the Consolidated Framework for Implementation Research. *PLOS One, 18*(4), e0278007. doi:https://doi.org/10.1371/journal.pone.0278007
13. Edemekong, P. F., Annamaraju, P., Afzal, M., & Haydei, M. J. (2024). Health Insurance Portability and Accountability Act (HIPAA) Compliance. *Treasure Island*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK500019/
14. Ekin, T., Leva, F., Ruggeri, F., & Soyer, R. (2013). Application of Bayesian Methods in Detection of Healthcare Fraud. *Chemical Engineering Transactions*, 151-156.
15. Field, R. I. (2017). Regulation of health care in the United States: complexity, confrontation and compromise. *An Inst Hig Med Trop, 16*(1), 61-70.
16. Glynn, E. H. (2022). Corruption in the health sector: A problem in need of a systems-thinking approach. *Frontiers in Public Health, 10*, 3389. doi:https://doi.org/10.3389/fpubh.2022.910073
17. Goel, R. K. (2020). Medical professionals and health care fraud: Do they aid or check abuse? *Managerial and Decision Economics, 41*(4), 520-528. doi:https://doi.org/10.1002/mde.3117
18. Gostin, L. O., & Katz, R. (2016). The International Health Regulations: The Governing Framework for Global Health Security. *The Milbank Quarterly, 94*(2), 264-313. doi:https://doi.org/10.1111/1468-0009.12186
19. Hasan, M. N., Arman, M., Bhuyain, M. M., Chowdhury, F., & Bathula, M. K. (2025). Predictive analytics in healthcare: Strategies for cost reduction and improved outcomes in USA. *International Journal of Innovative Research and Scientific Studies, 8*(8), 142-150.
20. Haynes, K., Agiro, A., Chen, X., Stephenson, J. J., Eshete, B., Sutphen, R., . . . Young, K. (2020). Developing Methods to Link Patient Records across Data Sets That Preserve Patient Privacy. *Patient-Centered Outcomes Research Institute*.
21. Ikono, R., Iroju, O., Olaleke, J., & Oyegoke, T. (2019). A Meta-Analysis of Fraud, Waste and Abuse Detection Methods in Healthcare. *Nigerian Journal of Technology, 38*(2), 490-502. doi:http://dx.doi.org/10.4314/njt.v38i2.28
22. Iqbal, M. S., Abd-Alrazaq, A., & Househ, M. (2022). Artificial Intelligence Solutions to Detect Fraud in Healthcare Settings: A Scoping Review. *Advances in Informatics, Management and Technology in Healthcare*, 20-23. doi:https://doi.org/10.3233/SHTI220649

23. Johnson, J. M., & Khoshgoftaar, T. M. (2023). Data-Centric AI for Healthcare Fraud Detection. *S N Computer Science, 4*(4), 389. doi:https://doi.org/10.1007/s42979-023-01809-x
24. Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2014). Using Data Mining to Detect Health Care Fraud and Abuse: A Review of Literature. *Global Journal of Health Science, 7*(1), 194-202. doi:https://doi.org/10.5539/gjhs.v7n1p194
25. Kumaraswamy, N., Markey, M. K., Ekin, T., Barner, J. C., & Rascati, K. (2022). Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead. *Persepctives in Health Information Management, 19*(1).
26. Leder-Luis, J., & Malani, A. (2025). The Economics of Healthcare Fraud. 1-43.
27. Lederman, L. (2021). The Fraud Triangle and Tax Evasion. *Digital Repository @ Maurer Law*, 1-56.
28. Mackey, T. K., Miyachi, K., Fung, D., Qian, S., & Short, J. (2020). Combating Health Care Fraud and Abuse: Conceptualization and Prototyping Study of a Blockchain Antifraud Framework. *Journal Medical Internet Research, 22*(9), e18623. doi:https://doi.org/10.2196/18623
29. Makandah, E. A., Aniebonam, E. E., Okpeseyi, S. B., & Waheed, O. O. (2025). AI-Driven Predictive Analytics for Fraud Detection in Healthcare: Developing a Proactive Approach to Identify and Prevent Fraudulent Activities. *International Journal of Innovative Science and Research Technology, 10*(1), 1521-1529. doi:https://doi.org/10.5281/zenodo.14769423
30. Marquart, J. W., & Thompson, R. A. (2024). Exploring relation fraud, murder,and the Fraud Triangle. *Journal of Economic Criminology, 4*, 100061. doi:https://doi.org/10.1016/j.jeconc.2024.100061
31. Massi, M. C., Ieva, F., & Lettieri, E. (2020). Data mining application to healthcare fraud detection: a two-step unsupervised clustering method for outlier detection with administrative databases. *BMC Medical Informatics and Decision Making, 20*, 160. doi:https://doi.org/10.1186/s12911-020-01143-9
32. Muhith, A., Wibowo, N. M., Widiastuti, Y., & Utari, W. (2019). Detection of Healthcare Fraud in The National Health Insurance Program Based on Cost Control. *Advances in Economics, Business and Management Research, 103*, 284-288. doi:https://doi.org/10.2991/TEAMS-19.2019.46
33. Munappy, A. R., Bosch, J., Olsson, H. H., Arpteg, A., & Brinne, B. (2022). Data management for production quality deep learning models: Challenges and solutions. *Journal of Systems and Software, 191*, 111359. doi:https://doi.org/10.1016/j.jss.2022.111359
34. Oikonomou, E., Carthey, J., Macrae, C., & Vincent, C. (2018). Patient safety regulation in the NHS: mapping the regulatory landscape of healthcare. *Bio-medical Journal*, e028663. doi:https://doi.org/10.1136/bmjopen-2018-028663
35. Omair, B., & Alturki, A. (2020). A Systematic Literature Review of Fraud Detection Metrics in Business Processes. *IEEE Access, 1*(1), 99. doi:https://doi.org/10.1109/ACCESS.2020.2971604
36. Parti, K. (2023). What is a capable guardian to older fraud victims? Comparison of younger and older victims' characteristics of online fraud utilizing routine activity theory. *Frontiers in Psychology, 14*(1), 1118741. doi:https://doi.org/10.3389/fpsyg.2023.1118741
37. Pegu, N., Seth, S., Ramakrishnan, S., & Jangili. (2025). Healthcare Predictive Modeling for Identifying Fraud in Medical Insurance Claims. *International Journal of Pharmaceutical Science, 3*(2), 1734-1744. doi:https://doi.org/10.5281/zenodo.14899939
38. Preez, A. D., Bhattacharya, S., Beling, P., & Bowen, E. (2025). Fraud Detection in Healthcare Claims Using Machine Learning: A Systematic Review. *Artificial Intelligence in Medicine, 160*, 103061. doi:https://doi.org/10.1016/j.artmed.2024.103061
39. Samonas, S. (2013). Insider fraud and routine activity theory. *LSE Research Online*, 1-32.
40. Sayem, M. A., Taslima, N., Sidhu, G. S., & Ferry, J. W. (2024). A Quantitative Analysis of Healthcare Fraud and Utilization of AI for Mitigation. *International Journal of Business and Management Sciences*, 13-36. doi:https://doi.org/10.55640/ijbms-04-07-03
41. Schweppenstedde, D., Hinrichs, S., Ogbu, U., Schneider, E. C., Kringos, D. S., Klazinga, N. S., . . . Nolte, E. (2014). Regulating Quality and Safety of Health and Social Care. *Rand Health Quarterly, 4*(1), 1.
42. Settipalli, L., & Gangadharan, G. R. (2021). Healthcare fraud detection using primitive sub peer group analysis. *Concurrency and Computation: Practice and Experience, 33*(23), 62-75. doi:https://doi.org/10.1002/cpe.6275

43. Stowell, N. F., Pacini, C., Wadlinger, N., Crain, J. M., & Schmidt, M. (2020). Investigating Healthcare Fraud: Its Scope, Applicable Laws, and Regulations. *William & Mary Businesss Law Review, 11*(2), 1-50.
44. Sujeewa, G. M., Ab Yajid, M. S., Khatibi, A., Azam, S. M., & Dharmaratne, I. (2018). The New Fraud Triangle Theory: Integrating Ethical Values of Employees. *International Journal of Business, Economics and Law, 16*(5), 52-57.
45. Sumalatha, M. R., & Prabha, M. (2019). Mediclaim Fraud Detection and Management Using Predictive Analytics. *International Conference on Computational Intelligence and Knowledge Economy*. doi:https://doi.org/10.1109/ICCIKE47802.2019.9004241
46. Szewczyk, T., Sinha, M. S., Gerling, J., Zhang, J. K., Mercier, P., & Mattei, T. A. (2024). Health Care Fraud and Abuse: Lessons From One of the Largest Scandals of the 21st Century in the Field of Spine Surgery. *Annals of Surgery Open, 5*(2), e452. doi:https://doi.org/10.1097/AS9.0000000000000452
47. Thaifur, A. Y., Maidin, M. A., Sidin, I. A., & Razak, A. (2021). How to detect healthcare fraud? "A systematic review". *Gaceta Sanitaria, 35*(2), 441-449. doi:https://doi.org/10.1016/j.gaceta.2021.07.022
48. Thomas, S. S., & Sheshassayee, A. (2017). Influence of Predictive Analysis in Helath Insurance Fraud Detection. *International Journal of Engineering Applied Sciences and Technology, 2*(7), 25-27.
49. Tickner, P., & Button, M. (2021). Deconstructing the Origins of Cressey's Fraud Triangle. *Journal of Financial Crime*, 1-12.
50. Xiao, F., Li, H. X., Wang, X. K., & Wang, J. (2025). Predictive analysis for healthcare fraud detection: Integration of probabilistic model and interpretable machine learning. *Journal of Information Science, 12*(24). doi:https://doi.org/10.1016/j.ins.2025.122499