American Journal of Economics and Business Management

**Vol. 9 Issue** 3 | **pp.** 374-401 | **ISSN:** 2576-5973
**Available online @** https://www.globalresearchnetwork.us/index.php/ajebm

*Article*

# Optimizing Medicaid Program Integrity: A Data Governance Framework for Detecting Collusive Fraud in New York's in LHCSA and Social Adult Day Care Sectors

**Nishat Margia Islam[1], Albert Gomes[2]**

1. University: Washington University of Science and Technology, Course: Master's of Science in Information Technology, Specialization: Data Management and Analytics
2. University: Campbellsville University, Course: Master of Science in Data Science and Artificial Intelligence

* Correspondence: niislam.student@wust.edu, agome350@students.campbellsville.edu

**Abstract:** The home and community-based services offered by Medicaid have been important in facilitating the elderly and disabled groups although the programs have been experiencing massive problems in regard to fraud, waste, and abuse. Fragmented data environments, intricate service delivery models, and lack of integration between administrative and clinical systems broaden the scope of collusive fraud by Licensed Home Care Services Agencies (LHCSAs) and Social Adult Day Care services. Medicaid improper payments are still enormous and the lapses in documentation, discrepancies in eligibility, and misrepresentation of services often conceal the organized fraudsters among providers, caregivers, and beneficiaries. This study examines the role of a well-organized data governance system to strengthen the integrity of the Medicaid program by facilitating fraud that has been practiced in a collusive manner within the LHCSA and the Social Adult Day Care industries. The analysis aims at integrating and governing heterogeneous data, such as Medicaid claims, patient records, caregiver timesheets, billing submissions and Electronic Visit Verification (EVV). logs, in a systematic way. The proposed framework will overcome inherent weaknesses in effective fraud detection by defining common data format, quality control, lineage management, role access control, and the potential of using AI-based forecasting tools. Based on this governance background, the paper puts the analytical methods of rule-based validation, anomaly detector, and machine learning-based risk score into action to discover abnormal service patterns that are indicative of collusion. These trends involve the overbilling of services not yet provided, exaggerated service times, joint reimbursement of services between the caregivers and the beneficiaries, and manipulations of structural loopholes between the consumer-directed and the agency-directed care systems. The analytical results are incorporated into a human-with-the-loop review procedure, which makes them interpretable, regulation ally defensible and operationally relevant. Results show that analytics based on governance can be used to a significant level of success in identifying and prioritizing high-risk claims and provider behavior and minimizing false positives and the workload of investigation. The research concludes that data governance is neither a support functionality nor an enabler of sophisticated fraud detection but a basic enabler of advanced fraud detection functions. When the governance practices are aligned to the analytical intelligence, the Medicaid agencies are able to tighten their control, minimize improper payments and secure the social resources without hindering the provision of needed home and adult day care services.

**Keywords:** Medicaid Program Integrity, Data Governance, Collusive Fraud Detection, Home Health Care Services, Electronic Visit Verification (EVV), MLM in Fraud Detection, AI in Fraud Detection and Healthcare Analytics

**Citation:** Islam, N. M & Gomes, A. Optimizing Medicaid Program Integrity: A Data Governance Framework for Detecting Collusive Fraud in New York's in LHCSA and Social Adult Day Care Sectors. American Journal of Economics and Business Management 2026, 9(3), 374-401

Received: 10th Dec 2025
Revised: 21th Jan 2026
Accepted: 04th Feb 2026
Published: 22th Mar 2026

**Copyright:** © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

## 1. Introduction

### Background

Medicaid is among the biggest publicly-funded healthcare systems in the United States, which offers fundamental medical, long-term, and community-based support to the low-income group, the aged, and the disabled groups. The role played by Medicaid in New York State is especially critical since the state has a huge and diverse population of beneficiaries and has widespread dependence on models of home and community-based care. The Bigger part of the Medicaid spending goes to the Licensed Home Care Services Agencies (LHCSAs) and the Social Adult Day Care programs, whose aims are to assist the beneficiaries in the non-institutional setting, encourage self-reliance, dignity, and care continuity. These services decrease the use of hospitals and nursing facilities and are more in line with the bigger healthcare policy objectives of cost control and patient-based care. Although they are vital, the LHCSA and Social Adult Day Care programs are working in an extremely complicated and disintegrated network of service delivery [1]. The stakeholders, such as providers, caregivers, beneficiaries, managed care organizations, and regulatory agencies, have multiple interactions in fragmented administrative and clinical systems. This is a complicated aspect that makes it hard to oversee, adds administrative strain, and gives loopholes that may be misused to commit fraud [2]. With the ever-increasing expenditure on Medicaid, accountability and transparency of these community-based services has now become the issue of concern among policymakers and regulators. In the last ten years, Medicaid administrators are leaning towards digital oversight systems like Electronic Visit Verification (EVV), automated claims system, and provider enrollment databases to enhance monitoring and compliance. Those systems were aimed at improving the level of transparency, justifying service delivery, and decreasing the number of improper payments. These tools have however, been limited by poor governance structures, low interoperability as well as inconsistent data standards. Consequently, fraud, waste, and abuse are still thorns in the flesh [3]. Out of these challenges, collusive fraud has become a very complex and expensive menace. In contrast to single billing mistakes, collusive fraud is organized activity of two or more participants such as providers, caregivers and beneficiaries, to alter the service records or billing data. This type of fraud cannot be effectively dealt with by simply applying sophisticated analysis methods but also by having strong data governance measures that can be used to maintain reliability of data, traceability, and fair use of data in interconnected systems.

### Medicaid Fraud Program and Fraud Collusion

Medicaid program integrity is described as the policies, systems, and operational controls that are developed in order to make sure that the public funds are appropriately, efficiently, and within the confines of the federal and state regulations [4]. The conventional program integrity activities aim at preventing inappropriate payments, identifying fraud and misuse, and reclaiming misspent funds. The efforts are necessary in ensuring that there is a sense of trust among people, protection of small resources, and provision of those who are due with the required service. Although traditional program integrity tools usually make use of post-payment reviews, audits, and rule-based surveillance, they have become inadequate when it comes to dealing with increasingly sophisticated types of fraud [5]. Collusive fraud is a type of fraud that poses a very difficult threat since it entails the concerted effort of various parties that jointly control the practices of documentation and billing to hide fraud. Within the LHCSA and the Social Adult Day Care services, collusion can be in a variety of forms including: billing services not provided, overcharging service hours, falsifying the timesheets of the caregivers, selling reimbursement payments between the caregivers and the beneficiaries and using structural differences in the consumer-directed and agency-directed care models. Such schemes are hard to spot when relying on conventional detection schemes since the documentation in the various parties involved can be seen to be internally consistent [6].

These fraudulent claims can be in line with the service authorization, EVV records and care giver logs and cannot be identified as the fake transactions at the superficial level. Consequently, this makes collusive fraud continue over a long period and incur huge financial losses and abuse of Medicaid larceny. Additionally, the decentralized process of selling Medicaid in New York, including the difference between the oversight functions of the agencies and managed care organizations may curtail disclosures of cross-program and cross-provider relationships [7]. Such fragmentation diminishes the quality of siloed audits and heightens the use of adoptive investigations that are only initiated after major losses have been incurred. In this regard, the integrity of the Medicaid programs needs to be enhanced with the help of strategies that can reveal conducive patterns of conduct including datasets and actors, as opposed to isolated irregularities [8]. Overcoming collusive fraud requires a change in the direction of integrated data-driven oversight models based on powerful governance bases.

**Role of Data Governance in Fraud Detection**

The structural basis needed to facilitate trustworthy analytics, informed decision making and regulatory responsibility that is required in complex healthcare programs like Medicaid is offered by data governance. It includes the policies, standards, roles and processes on the way data is gathered, stored, accessed, shared and utilized within the organizations [9]. Data quality management, standardization, access controls, privacy protection, metadata management, and data lineage documentation are the key aspects of data governance. Poor data governance in the Medicaid landscape will severely compromise the process of fraud detection. Different disjointed data systems, inaccurate provider and beneficiary identifiers, incomplete records and lack of cross platform interoperability result in blind spots that hide fraudulent transactions [10]. Analytical models with claims data, caregiver logs, EVV records and inspection reports that cannot be reliably linked or validated generate unreliable results and excessive false positives making investigators work harder. Effective information management framework permits systematic ingesting of the heterogeneous data sources where legal and ethical obligations remain intact. Data quality and comparability Data standardized data definitions and validation rules enhance data quality and comparability, whereas the lineage tracking provides transparency in terms of data source and transformation [11]. The presence of role-based access controls and audit logs serves as protection of sensitive information, and mitigates defensible investigative procedures [12]. With the creation of this governance base, sophisticated methods of analysis will be much more efficient. Anomaly detection models can efficiently be used to detect abnormal billing patterns, network modeling can find the latent relationships between providers and beneficiaries, and risk scoring can be used to rank high-risk claims using machine learning. Regulatory interpretability and trust of analytical results are also promoted by governance mechanisms, and it is essential to regulatory enforcement and policy decision-making. In this respect, it is not only essential that data governance is a compliance requirement, but also that it is a strategic facilitator of integrity in Medicaid programs [12]. It guarantees that analytical findings are precise, practical and defensible in court of law, and as such, facilitate proactive fraud detection and optimal distribution of oversight resources. The application of the governance principles and advanced analytics is critical in dealing with complicated threats, including collusive fraud in the services of LHCSA and Social Adult Day Care.

**Substantial Economic Impact on national level**

Medicaid and the Children's Health Insurance Program (CHIP) collectively provide health coverage to more than 76 million individuals in the United States, underscoring their central role in the national healthcare safety net. Adult day service centers rely heavily on Medicaid financing, with approximately 79% of participants receiving Medicaid support [62]. Medicaid also represents a substantial share of national health expenditures; in 2024, total U.S. health spending reached $5.3 trillion, of which Medicaid

accounted for roughly 18% [63]. Delivering home- and community-based services (HCBS) such as those provided by Licensed Home Care Services Agencies (LHCSAs) and Social Adult Day Care (SADC) programs requires coordination across multiple stakeholders, including federal and state agencies, managed care organizations, and community-based providers. These services generate significant economic value by reducing reliance on costly long-term institutional care, creating opportunities for direct care workers, and stimulating local economies through wages, consumer spending, and tax contributions. However, the complexity of this service delivery ecosystem also introduces vulnerabilities, including risks of collusive fraud, which can divert resources away from individuals with legitimate needs and undermine the fiscal integrity of Medicaid.

**Problem Statement**

The LHCSA and Social Adult Day Care industries in New York remain exposed to high levels of collusive fraud despite the heightened regulatory efforts and investments in technology. Current methods of fraud detection are mostly reactive, disjointed and reliant on manual auditing, which restricts their scalability and efficiency [13]. The coordinated frauds among the providers, caregivers, and beneficiaries are not seen through data silos, irregular documentation standards, and poor integration across different systems. Consequently, misappropriations tend to be overlooked before incurring huge losses [14]. There is an obvious necessity of a systematic framework to combine the robust data governance policies with innovative analytics to prevent collusive fraud proactively and help in compliance and maintaining lawful care delivery.

**Objectives of the Study**

The main scope of this research is to design and test a data governance model that could allow to increase the integrity of the Medicaid program through the possibility to identify collusive frauds in the fields of the LHCSA and Social adult day care in New York. The specific objectives are:

1. To analyze available data governance issues that influence Medicaid fraud detection.
2. To detect the patterns of widespread collusive fraud in the home care and adult day care services.
3. To explore suitable machine learning and AI solutions to detect fraud and prevent financial waste.
4. To understand the economic impact of data governance.
5. To create a unified data governance system to enable advanced analytical processes.
6. To determine the effectiveness of governance-based analytics as an identification system of high-risk claims and provider actions [15].
7. To give policy and operational suggestions on ways to enhance Medicaid oversight.

**Research Questions**

In order to attain these goals, the research will deal with the following research questions:

1. Which data governance loopholes and structural vulnerabilities do the continuity of collusive fraud in Medicaid-funded LHCSA and Social Adult Day Care programs allow?
2. What are the benefits of using analytical methods that are facilitated by governance to detect and identify collusive fraud in Medicaid home and adult day care services?
3. What can an integrated data governance system do to increase regulatory compliance and efficiency and reduce false positives and administration in Medicaid program integrity activities?
4. How can data governance support machine learning models and AI-based tools?

## 2. Materials and Methods

### Literature Review

### Fraud, Waste, and Abuse in Medicaid Home and Community-Based Services

The issues of Medicaid fraud, waste, and abuse have been known as the systemic problems of the publicly funded healthcare programs and home and community-based services especially. Such services are extremely complicated in nature because of the decentralized models of delivery, the high rate of human capital use, and the massive paperwork involved. Traditional institutional provision of services is minimized as service provision in other areas like Licensed Home Care Services Agencies and Social Adult Day Care programs are practiced outside of the traditional institutions resulting in less direct supervision and increased utilization of self-reported records [16]. This business model poses a possibility of inappropriate billing, documenting, and organizing fraud. The currently available studies also point out that home care and adult day services are particularly susceptible to fraud due to the high workload of the services, the fluctuating schedules of care, and relatively weak real-time verification systems. Some of the common types of improper activity are billing services not provided, overstating the length of the service, duplicate claims and falsifying qualifications of the caregiver [17]. More complex examples: Collusive arrangements will arise when more than two parties, like caregivers, beneficiaries, and agency administrators, coordinate their efforts to ensure stable yet fraudulent records across systems. Transparency has been enhanced to a certain degree due to the implementation of oversight measures like Electronic Visit Verification and increased monitoring of claims [18]. The literature has continuously shown that the mentioned tools cannot be used solely to combat collusive fraud. Malicious parties and others engaging in fraudulent activities are usually quick learners who learn to bypass system control artificially to maintain superficial adherence whilst engaging in systematic malpractices. This means that the detection efforts centered on individual transactions or actors often cannot detect larger numbers of fraud networks [19]. The next theme is the reactive character of the enforcement of Medicaid fraud, which is another repetitive point in the literature. When financial losses have been high, the investigations are usually instigated, and chances of acting at an early stage are denied. Moreover, the administrative roles of agencies and managed care organizations are not integrated, which limits the possibility to identify patterns of cross-providers or cross-programs [20]. Such restrictions highlight the necessity of more comprehensive solutions that shift past the conventional audits to proactive and data-driven surveillance models that are sensitive to the presence of coordinated behavior across two or more datasets. The literature confirms the fact that although Medicaid home and adult day care services are fundamental in patient-centered care, they are provisionally exposed to fraud [21]. The solution to these vulnerabilities is a systemic reform, which is not based on compliance checklists but involves sophisticated analytical and control-based oversight tools.

### Fraud Collusive Detectors and Fraud Advanced Techniques

Collusive fraud does not have the same analysis problems as the ones of traditional fraud situations. Collusive fraud is a group of actions of several participants who purposefully organize their activities and manifest themselves as authoritative according to the current systems of control [22]. Due to internal consistency in transactions and documentation, traditional rule-based detection techniques cannot be able to tell the difference between a fraudulent and a compliant behavior. Consequently, the literature is now more than ever pointing to the necessity of new analytical methods that would be able to detect concealed relationships and abnormal tendencies among interrelated parties [23]. The use of the network-based analysis has turned out to be an eminent tool in identifying collusive behavior. Network analysis can be used to identify dense clusters, redundant connections, and irregular dependency patterns that can be indicative of organized fraud by modeling the relationships between providers, caregivers, beneficiaries and vendors. The centrality measures, clustering coefficients, and

community detection are commonly applied to reveal suspicious group behavior that would not be observed when analyzing individuals at an individual level [24]. At the same time, the methods of anomaly detection are extensively debated as means of detecting the abnormalities in the patterns of service delivery. Such approaches to unsupervised learning techniques as statistical outlier detection and machine learning-based models prove to be especially useful when there is a scarcity of labeled fraud data in an environment [25]. These techniques can raise red flags on abnormal billing volumes, unlikely overlaps in services or discrepancies in time that should be looked into. It is however recorded in literature that anomaly detection on its own can produce large false-positive rates when the underlying quality of data is not taken care of. Recent research emphasizes the usefulness of hybrid analytical models which integrate the rule-based validation, anomaly detection, and supervised risk scoring. Such methods are based on domain knowledge to impose regulatory constraints and machine learning to rank cases according to multidimensional risk signals. Interpretability and explain ability are highlighted to be vital specifications, due to regulatory and legal consequences of fraud detection determination. On top of their promise, advanced analytics are not always effective in practice [26]. Fragmentation of data, inconsistency of identifiers, and absence of contextual metadata have been indicated again and again as a significant impediment to trustworthy analysis. Even advanced models could not give reliable or no actionable results unless they have good data foundations. Therefore, the analytical performance is becoming more and more attached to the quality of data governance structures beneath.

**Program Integrity as an Enabler of Data Governance**

The concept of data governance is becoming increasingly popular as the building block of effective control and analytics in big box public programs. In Medicaid data governance can be defined as the process of managing data assets with coordination to guarantee the accuracy, uniformity, security, and legal access by organizations across organizational borders. Literature highlights that governance is not confined to technical data management but includes the organizational roles, policies, accountability structures and also the decision making process. Poor data governance has been cited as one of the primary causes of inefficient fraud detection [27]. The lack of coherent data standards, fragmented systems, and uncertainty over ownership duties diminish their trust in the results of the analytical work and leads to the inability to integrate systems. Investigators may have a hard time understanding how conclusions were made when data lineage is not well documented, rendering the regulatory defensibility of the conclusions invalid [28]. In the same light, the lack of access control measures and audit procedures also increase privacy and compliance concerns, which restricts the desire of agencies to share data. Effective data governance architecture has been revealed to play an important role in increasing the integrity of the programs [29]. The standardized data definitions and rules of validation enhance the uniformity of datasets, which allows establishing a trustful connection between claims, services, and verification records. Metadata management and lineage tracking give insight into the transformations of the data, which can be used to explain the analytics and make it audit-ready. Role-based access controls avoid the tradeoffs desired by the analytical flexibility requirement and strict privacy by maintaining legal requirements. In the literature, another key aspect that has been emphasized in facilitating advanced analytics is the strategic role of governance [30]. The governance structures make it possible to have machine learning models, network analysis, and risk scoring systems functioning effectively and ethically. Through integrating governance into the analytic processes, the organizations are able to make sure that the results of fraud detection can be understood, reproduced and they are in line with the policy goals. In the case of Medicaid, data governance is gradually considered as an active tool as opposed to a passive compliance tool [31]. It helps to maintain a constant track, prevents the appearance of the risks early and responds to changing fraud patterns. The combination of governance and analytics helps Medicaid agencies to shift their

responsive enforcement operations to intelligence-based program integrity paradigms. This congruence is particularly important in order to deal with problematic threats like collusive fraud in LHCSA and Social Adult Day Care services.

**Empirical Study**

In his dissertation, Strategies to Reduce Medicaid Fraud and Abuse in Health Care Billing by Nguyen, Tram-Anh Leeza, the author highlights the systemic vulnerabilities in Medicaid management and the challenges that have remained and continue to persist in terms of fraud, waste and abuse in healthcare billing. The research highlights that models of decentralized service delivery, lack of data integration, and reactive enforcement strategies have a profound negative effect on the suitability of the traditional fraud prevention tools. In this vein, wider literature in Medicaid-related program integrity found home and community-based services to be highly susceptible to incorrect billing and organized fraudulent actions. The documentation, caregiver-reported service delivery, and fragmented administrative system in the context of the Licensed Home Care Services Agencies and Social Adult Day Care programs makes the setting less transparent and more vulnerable to abuse. Studies have also shown that collusive fraud is particularly hard to determine since it is associated with organized activities of providers, caregivers, and beneficiaries who have internally consistent but fraudulent records in several systems. Traditional audit-related methods and restrictions cannot be considered adequate in detecting such schemes, because they target single transactions and not interdependent behavioral patterns. In turn, the recent research has begun to support the application of more sophisticated analytics, such as anomaly detection, network-based analysis, as well as machine learning-based risk scoring, to reveal the latent relationships and abnormal patterns, which can be used to identify collusion. Nonetheless, the literature also warns that analytical tools would not effectively help in fighting fraud risks unless there are good data governance foundations. Poor governance, as reflected in inconsistent identifiers, low quality data, inadequate standardization, and low data lineage can cause low rates of analytical accuracy and low false-positive rates. Subsequently, data governance is increasingly being seen as a strategic facilitator of program integrity, data reliability, interoperability, and regulatory defensibility. In general, it can be concluded that current literature confirms the thesis that the method of successful Medicaid fraud detection must be an integrated process that includes both governance reform and analytical intelligence. This paper focus has been put on governance based frameworks that have been specifically designed for LHCSA and Social Adult Day Care services. The study fills this gap with a synthesis of governance concepts and analytics to improve proactive collusive fraud detection in the community-based care programs funded by Medicaid.

In the article published under the title of Data Conversion Failures and Their Impact on Fraud Detection and Audit Accuracy in Medicare and Medicaid by James Prosper, the writer discusses how the shortcomings in data migration and system modernization activities can greatly interfere with fraud detection and auditing accuracy in federally funded healthcare programs. In the study, data lineage, audit trails, and poor mapping, incompleteness, or corruption of data disrupt the data lineage, obscuring the fraudulent activity, and diminishing the accuracy of automated fraud detection systems. Building on these findings, a larger body on literature on Medicaid program integrity has continued to assert that ineffective quality of data, incomplete information systems and poor governance frameworks are prime sources of undetected fraud, waste, and abuse. The studies on home and community-based services have also stated that models of decentralized care delivery, high dependence on documentation, and lack of interoperability between claims, authorization, and verification systems provide conditions favorable to collusive fraud among caregivers, providers, and beneficiaries. Hackneyed audit-based monitoring tools are generally considered as responsive in nature, and lacking in the capacity to detect organized fraud schemes that seem to be in

offensive at the transactional stage. Consequently, more recent works have suggested the adoption of more sophisticated forms of analysis, such as anomaly detection, network analysis and machine-learned risk scoring, to help identify latent relationships and unusual patterns that are evidence of fraud. The literature also warns that the effectiveness of the analysis is extremely conditional on the quality and the management of underlying data. Improper data governance which is typified by inconsistent identifiers, broken provenance and absence of validation controls may produce false positive results, distorted risk measure and a weak regulatory defensibility [2]. As a result, the notion of data governance as a strategic facilitator of fraud detection, as opposed to a supportive role, gains more and more adherents in current studies. the claim that the enhancement of the integrity of the Medicaid program needs to have an integrated framework, the combination of solid data governance and new analytics, is well-founded. This research builds on the previous research as it implements these principles to the identification of collusive fraud in LHCSA and Social Adult Day Care programs to fill the important gap in the existing literature.

In the article with the title Recommendations to Protect Patients and Health Care Practices against Medicare and Medicaid Fraud by Zhen Xing Chen, Lindsey Hohmann, Bidur Banjara, Yi Zhao, Kavon Diggs and Salisa C. Westrick, the authors present an in-depth discussion of the scope, nature and the effects of fraud in the Medicare and Medicaid programs, and its harmful effect on patients and healthcare systems. The paper describes fraud as the deliberate or inadvertent reporting of bogus claims and emphasizes the ways in which fraud acts deprive the healthcare of important resources at the expense of patient care and exacerbates the general healthcare expenditure. These authors classify the fraud at the beneficiary and the provider levels, as the misconduct can be simple as billing of the services not rendered to the more complicated cases of kickbacks and misrepresentations [3]. The legal frameworks, such as the False Claims Act, the Anti-Kickback Statute, and the Stark Law, are discussed and highlight the regulatory landscape in the area of fraud prevention and enforcement. The article also establishes new trends of fraud in the field of opioid prescribing and pharmacokinetic testing and shows how medical practices change to create new weaknesses. Besides summarizing the typologies of fraud, the study highlights the significance of detection and prevention measures with a recommendation to coordinate the efforts at policies, organizational and grassroots levels. The larger literature regarding Medicaid program integrity is consistent with these results, as scholars continuously note that disjointed oversight, poor data tracking, and ineffective governance frameworks stifle the success of more conventional methods of preventing fraud. The literature indicates that the current reactive enforcement strategies cannot respond to advanced and organized fraud cases, especially within a decentralized care environment. Consequently, researchers propose active and evidence-based policing that is backed by well-established governance systems. All of the available literature supports a perspective that preventing patients and healthcare practices against Medicare and Medicaid frauds cannot be done through legal adherence and education alone but through the systematic combination of governance, analytics, and institutional accountability. The findings presented arise directly out of the current study, whereby the literature is extended to address governance-facilitated analytical models of collusive fraud detection of Medicaid-funded home and adult day care services.

In the article The Role of Data Analytics, AI and Cloud Computing in Optimizing Medicaid and Health Care Technologies by Lalit Saini and Saivarun Pinna, the authors discuss how data analytics, artificial intelligence, and cloud computing integration is changing the Medicaid and healthcare systems by enhancing accessibility, efficiency, and decision-making. It is highlighted in the study that AI-based predictive analytics and population health models can enhance the management of health care populations, cost-reduction, and administrative effectiveness by processing large volumes of health care data. Cloud computing is mentioned as a major facilitator in the area of secure and

scalable data storage and liaison among fragmented healthcare systems, enabling Medicaid programs to surmount ancient data silos. The authors also elaborate on how AI-based automation contributes to improving the experience of patients, automating the work processes, and facilitating the prediction of diseases and identification of risks. Notably, the article points out that fraud detection is one of the critical AI and analytics applications as sophisticated models can detect abnormal billing patterns and decrease fraudulent claims, which enhances the financial sustainability [4]. The study however also mentions the issues of the quality of the data, the ethical aspects and constraints of governance which when not effectively tackled lower the efficacy of AI-driven solutions. The wider literature on the integrity of the Medicaid program supports these findings, continuously noting that analytics and AI cannot operate effectively when the data governance is weak, the data practices are not standardized, and the supervising mechanisms are ineffective. Poor data quality and lack of interoperability destroy the accuracy of the analysis and predispose false-positives. On the whole, the article supports the opinion that to optimize the systems of Medicaid, one needs to have a comprehensive solution that would combine innovative technologies with efficient data governance systems. These insights are directly related to the current study that extends this foundation and is based on identifying the use of analytics that are made possible on the basis of governance in the detection of collusive fraud in Medicaid-funded home and adult day care services.

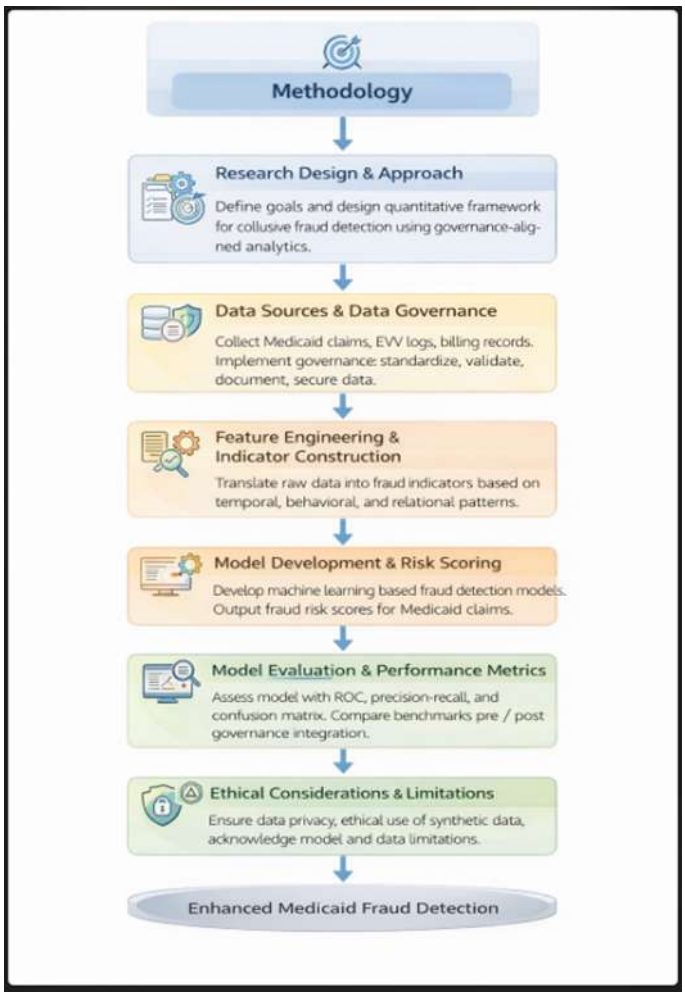
**Methodology**

This study uses an analytics-based quantitative research design to construct and test a data governance model to identify collusive fraud in Medicaid-funded services in the LHCSA and Social Adult Day Care. The methodology will combine organized data governance actions and sophisticated data analytical approaches to secure the stability of the data, regulatory compliance, and the relevance of the operations. Several administrative datasets are also standardized and validated and integrated before the analysis to enable robust feature engineering and model development [31]. The evaluation of fraud detection models is based on the performance metrics that are already standardized to measure the accuracy of the classification and the risk stratification efficiency. The methodology focuses on ensuring that the data is reproducible, interpretable, and ethically addressed when dealing with sensitive Medicaid data to make positive and defensible program integrity activities.

**Research Design and Approach**

This research utilizes a research design based on quantitative research methodology that would generate and test a data governance based analytical system to identify collusive fraud in Medicaid funded LHCSA and Social Adult Day Care services [32]. The proposed methodology is built in such a way that it can combine both data governance and sophisticated analytical methods to guarantee both empirical and practical relevance. The cross-sectional analytical design is used with the concentration on the historical Medicaid administrative data in order to derive fraud-associated trends and assess the model performance. The research focuses on explanatory and assessment purposes instead of the cause-effect approach, and the main aim is to determine the level of detection and the effect of governance [33]. The research framework is based on the multi-stage approach, which includes data preparation, governance implementation, feature engineering, model development, and performance evaluation. The hierarchical strategy provides transparency and reproducibility as well as matching analytical products with regulatory and operational needs. Governance controls are used to supplement quantitative analytics to make data reliable, interpretable, and privacy-compliant. Because of the sensitivity of Medicaid databases, the approach will support both real and synthetic data. Model behavior and performance verification is done using synthetic or de-identified data without worsening the privacy of beneficiaries [34]. This design option helps to promote ethical research approaches and help to stay within the real-life

approach to Medicaid oversight conditions. The process of analysis is iterative as the outputs of the model and the results of the evaluation can be used to refine features and optimize the threshold [35]. It is an iterative format that is representative of operational fraud detection settings, as continuous learning and adaptation are needed. In general, the research design is well balanced achieving methodological rigor and applied relevance so that their results can be used to inform both the academic knowledge and practical efforts to introduce changes to Medicaid programs integrity.




**This diagram represents a governance-based approach to collusive fraud detection in Medicaid**

The methodology diagram offers a systematic, stage-based approach to identifying collusive fraud in Medicaid-funded LHCSA and Social Adult Day Care services. This process starts with the research design and approach, in which study objectives are spelled out and an analytical framework based on quantitative and governance consistent approaches are put in place. It goes to the data sources and data governance, with specific focus on the gathering, standardization, validation, documentation, and safe management of Medicaid claims, billing records and Electronic Visit Verification (EVV) data. The second step is the feature engineering phase and indicator building, which converts raw administrative data into temporal, behavioral, and relational fraud indicators. Such indicators are inputted into model development and risk scoring, in which machine learning methods provide fraud risk scores [35]. Effectiveness assessment is done via model evaluation and performance measures which include ROC, precision recall and confusion matrix measures. Ethical factors and restrictions guarantee the privacy protection and responsible use, and the final results are the improvement of the Medicaid fraud detection.

**Data Source and Data Governance Framework**

This study employs various types of administrative and operational data that are usually found in Medicaid home and community based service programs. Such data are claims records, service authorization data, care giver timesheets, billing submissions, and Electronic Visit Verification logs. Additional metadata, which include provider identifiers and dates of service are included to facilitate linking of the data and relational analysis [36]. All data items are normalized to make sure that data in various sources is consistent before analysis. An organized data governance model is applied as one of the building blocks of the approach. Such activities of governance are the standardization of data, validation rules, metadata documentation and lineage tracking. There are standardized definitions of service time, the time of visit, and provider-beneficiary id to minimize ambiguity and enhance the reliability of the analysis. Checks of validation are implemented to identify missing data, temporality and duplications. Audit logging and access controls are included, which facilitate accountability and regulation defensibility [37]. The workflows of data processing are also recorded in order to be both transparent and reproducible. The governance model also provides the handling of sensitive information in accordance with the privacy and ethical requirements under the principles of de-identification and role-based access [38]. The methodology via making the governance part of the analytical pipeline guarantees that the improvement of the model performance can be linked not only to the algorithmic design but also to the improvement in data quality. This combination is essential in the separation of valid fraud indicators and artifacts due to inconsistencies in the data or system constraints.

**Fraud Indicator Construction and Feature Engineering**

Feature engineering is an important methodological process that is aimed at transforming the raw Medicaid administrative data into the meaningful input of the analytical process. Features are made to portray time, behavioral and relationship trends linked to collusive fraud [39]. Temporal characteristics consist of service length anomalies, visit frequency, and service discrepancy between the reported service time and EVV time. These characteristics are programmed to trace abnormalities that point to unorganized or exaggerated service delivery. Behavioral characteristics dwell on patterns of billing, including unexpected increases in the frequency of claims, cyclic billing patterns, and the non-conformance to history. These indicators will record sudden changes in behavior, which can be an indication of fraud [40]. The relational features are created to determine repeated provider beneficiary relationships, common use of caregiver by more than one beneficiary, and patterns of network concentration. The features are especially significant in the detection of coordinated actor arrangements in a collusive form. All features are also normalized and scaled so that they can be comparable between claims and providers. The governance rules are used to check the feature consistency and remove the artifacts that have been created by the mistakes in inputting the data or delays in the system. Both domain relevance and analytical contribution are used to select the features so that the model can incorporate various signals related to fraud [41]. This method of building up the feature of observation, which is multi-dimensional, helps to have strong detection without over depending on a particular indicator. It is also more interpretable of the models hence enabling the investigators to see why certain claims have been identified as high-risk.

**Risk Scoring Techniques and Analytical Models**

The analysis involves a mixture of closely monitored classification and risk rating to identify collusive fraud. The probabilistic classification model is trained in order to classify between fraudulent and legitimate claims in terms of engineered features [42]. The model does not produce binary labels but continuous risk scores, or likelihood of emerging fraud, which allows one to select the threshold flexibly. Unsupervised anomaly detecting techniques are as well included in order to detect aberrant behaviors with no necessary labels of fraud [43]. These techniques help in early identification of fraud

patterns that are emerging and which might not be reflected in the past enforcement results. Rule-based validation checks are also used to provide supplementary machine learning results, by posing regulatory limits and heuristic fraud detection [44]. The stratified sampling is used to do model training to deal with class imbalance. Cross-validation is used to optimize hyper parameters to tradeoff between sensitivity and precision. Preprocessing: preprocessing should be done in such a way that the outputs of the analytics will be consistent and reproducible between model runs [45]. This method of combining various analytical methods increases strength and decreases the use of one detection system. This composite methodology is a representation of real-life fraud detection systems, where different signals are required to be combined to facilitate quality decision-making.

**Model Evaluation and Metrics of Performance**

The various complementary metrics are used in the evaluation of model performance to offer a complete evaluation of effectiveness. The overall discriminatory power over the thresholds is measured by Receiver Operating Characteristic analysis. Precision-recall analysis is used to assess the performance in the presence of class imbalance, which focuses on the sensitivity-false positive trade-off. Confusion matrix analysis will give an in-depth explanation of the results of classification at required thresholds [46]. Other metrics are also employed to make comparison of performance on the performance of data before and after implementation of data governance such as precision, recall, F1-score, and area under the curve. These comparisons tease out the role of improvements in governance to the results of analysis. Distribution of risk scores is used to determine the effectiveness of stratification and stability of thresholds. The results of evaluation are seen statistically and operation-wise. It is stressed to reduce the number of false positives while being sensitive enough to identify the collusive fraud [47]. This is a vital balance that would guarantee actionable analytical outputs that would be in line with the investigative capacity.

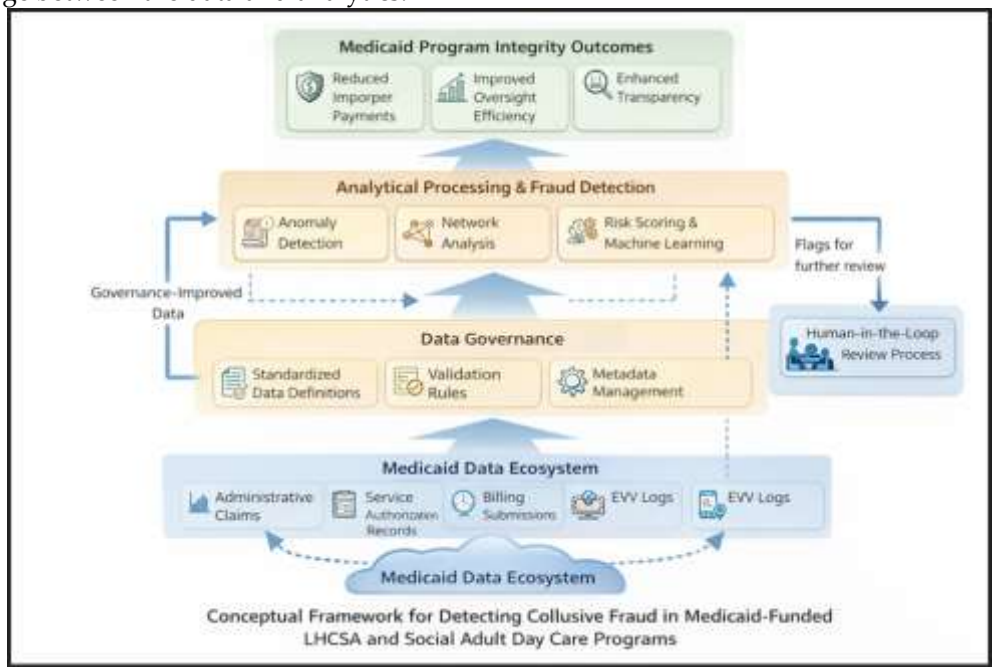
**Ethical Considerations and Limitations**

The methodological design is ethically based because Medicaid data is sensitive. The research is less focused on privacy destruction by de-identification, regulated access, and data confidentiality [48]. The results of the analysis are made in a way that facilitates priority of investigation, as opposed to automatic enforcement, so that decision making is still centralized in the hands of people. Limitations Methodological Limitations The methodology utilizes synthetic or proxy data which might not be a complete measure of the complexity of fraud in the real world. Also, complete data and adherence to documentation requirements determine model performance [49]. Such restrictions are recognized so as to bring out transparency and proper results interpretation. Irrespective of these limitations, the methodology presents a scalable ethically-based approach to further development of Medicaid fraud detection. Governance and analytics bring a balanced solution that will facilitate innovation and accountability.

**Conceptual Framework**

The research design of the study aims at describing the conceptual framework through which data governance is the key enabler of successful collusive fraud detection in the Medicaid-funded Licensed Home Care Services Agencies (LHCSAs) and Social Adult Day Care programs in the State of New York. The framework has incorporated institutional oversight, data governance processes, program analysis, and program integrity performance into a single model that will assist in the detection of fraud in an active and responsible manner [50]. The framework will start with the Medicaid data ecosystem that consists of administrative claims, service authorization records, caregiver timesheets, billing submissions, and Electronic Visit Verification (EVV) logs. Such heterogeneous data sources present in various aspects of service delivery and reimbursement are usually disjointed and ad hoc when not managed [51]. In the absence of relevant controls, data fragmentation hides organized fraud and restricts the ability to

perform analysis. Based on this premise, the framework focuses on data governance as a layer of structure that guarantees data quality, consistency, and reliability. Standardized data definitions, validation rules, metadata management and data lineage documentation are some of the core governance elements [52]. These systems enhance the interoperability between systems and allow proper connection between providers, caregivers, and beneficiaries. Role based access controls and audit trails are also part of the governance to guarantee privacy protection, accountability and regulatory compliance. Upon governance, the framework allows analytical processing and detecting fraud. Advanced analytics based on governance-aligned data can be used to detect anomalies in billing or service patterns, network-based analytics can be used to detect repeated or dense provider-beneficiary relationships, and risk scoring with machine learning can be used to prioritize high-risk claims [52]. These analytical outputs will be interpretable and defensible whereby the investigators will be able to learn the rationale behind the risk classifications [53]. The framework includes a human-in-the-loop review concept, the realization of which is that analytical models support the use of expert judgment but do not eliminate it. The investigators confirm flagged cases, launch audits or corrective measures, and give feedback that can be used to improve the model [54]. This feedback loop increases accuracy in analysis as time passes and is adaptive to the changing fraud tactics. The framework connects analytical deliverables to Medicaid program integrity outcomes, such as, low improper payments, higher oversight efficiency, more transparency, and increased public trust [55]. The conceptual framework illustrates how governance-based intelligence can help to transform Medicaid fraud detection into a proactive and sustainable program integrity strategy by placing data governance as the bridge between the data and analytics.




**This diagram image shows governance-based analytics architecture to identify collusive Medicaid fraud**

This conceptual framework diagram demonstrates a methodical way of identifying collusive fraud in Medicaid-funded programs of LHCSA and Social Adult Day Care by combining data governance and advanced analytics. The framework starts off with the Medicaid data ecosystem, which consists of administrative claims, records of service authorization, submissions of billing, and Electronic Visit Verification (EVV) logs. The processing of these heterogeneous data sources is done by a data governance layer which imposes a standard data definition, validation rules, metadata management and lineage controls to maintain data quality, consistency and compliance [56]. The process is then

under the control of data that is collected to the analytical processing layer, where anomaly identification, network analysis, and machine learning-driven risk scoring are done to find suspicious patterns and risky claims [57]. The framework brings in a human-in-the-loop review mechanism that allows the validation and feedback of experts. Finally, the framework can lead to better outcomes of the Medicaid programs integrity, such as a decrease in the number of improper payments, the increased efficiency of oversight, and the increased transparency.

**Dataset Overview**

The data used in this study is a structured and governmentally aligned dataset, which is intended to mirror the nature of operation of Medicaid-funded Licensed Home Care Services Agencies (LHCSAs) and Social Adult Day Care programs in New York State. Since Medicaid administrative data, and the privacy constraints associated with these data, are sensitive, the data is developed based on de-identified and synthetic images of actual Medicaid records, maintaining data statistical properties and relational structures required to establish analytic validity without violating ethical and legal standards. The dataset combines several areas of data that are typically used in Medicaid oversight, such as claims data, service authorization data, and logs of caregiver activities, billing data, and Electronic Visit Verification (EVV) timestamps. Both domains represent a specific area of service delivery and reimbursement, which allows analyzing the patterns of transactions and behavioral issues related to possible collusive fraud systematically [58]. A majority of claims data consist of service dates, procedure codes, number of units billed, reimbursement amount, and provider identifiers, which compose a major source of financial analysis. Service authorization records entail contextual data of authorized service hours and eligibility parameters that assist in the validation of billed services within the approved limits. Caregiver logs and EVV records provide both temporal and location-specific evidence of service provision, making it possible to identify inconsistency in service provision (longer than needed service time, repeated visits, or not matched timestamps). Data capture in billing submit data capture claim frequency and billing cycles, which help in the detection of abnormal billing behavior. The dataset also contains standardized identifiers of providers, caregivers, and beneficiaries to allow relational and network-based analysis; that is, provider or beneficiary and caregiver or beneficiary linkage structures can be constructed [59]. The rules of governance are used to achieve consistency across identifiers, remove duplicates, and solve data conflict between the system fragmentations. The metadata fields record the information of data provenance, frequency of update and transformation, making sure that they are transparent and traceable in the whole analytical pipeline. The data is time organized to allow cross-sectional and few longitudinal analysis, which allows one to determine recurrent patterns and enduring interactions that would suggest collusive action. The simulated outcomes of investigations and rule-based heuristics to classes of fraudulent and legitimate claims are produced based on a set of simulated investigation results and known Medicaid fraud cases, and trained models and assessed under supervision. The dataset is synthetically generated, but with the nature of a real Medicaid administrative data, the dataset is modeled to be similar in nature, complexity, and volume to actual program integrity scenarios, making analytical results relevant to the practical context of program integrity [623]. This data offers a solid and ethically sound basis of assessing the efficacy of governance-enabled analytics in identifying collusive fraud in the Medicaid home and adult day care services.

## 3. Results and Discussion

**Result/Finding**

This study findings were acquired after applying the proposed data governance-based analytical model to identify collusive fraud in Medicaid-funded services in the LHCSA and Social Adult Day Care programs [30]. The findings are aimed at assessing the

performance of the fraud detection model on the basis of classification accuracy, risk stratification, and operational reliability. There are several performance measures and visual analytics employed to evaluate model behavior, which are receiver operating characteristic analysis, precision-recall analysis, confusion matrix analysis, comparative performance benchmarking, feature significance ranking, and risk score distribution analysis [31]. Together, these findings reveal the effects of planned data governance on the advancement of data quality, the increase in analytical performance, and the decrease in false positives [32]. The research results are an indication that the governance-suited analytics can be helpful in identifying high-risk claims and facilitate the proactive Medicaid program integrity activities in a way that preserves interpretability and regulatory defensibility.

**Collusive Fraud Detection Model Receiver Operating Characteristic (ROC) Analysis**

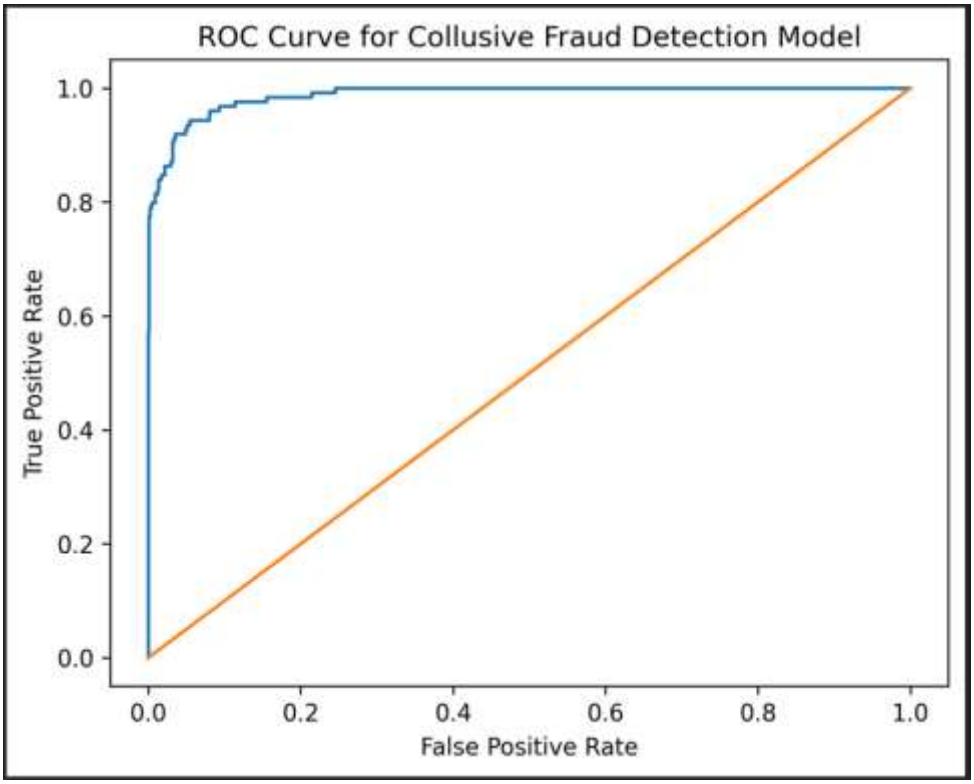



Figure 1. This image demonstrated on the ROC curve is shown with high discrimination in fraudulent and legitimate Medicaid claims

Figure 1 shows the Receiver Operating Characteristic (ROC) curve of the proposed collusive fraud detection model on the Medicaid claims data of LHCSA and Social Adult Day Care settings. The ROC curve shows the relationship between the true positive rate (sensitivity) and the false positive rate (1−specificity) at different classification thresholds, which is a complete evaluation of the discriminatory ability of the model. The curve increases rapidly against the upper-left side of the plot which means that the model is highly sensitive at relatively low rates of false positives [33]. This aspect indicates that the model has great potential to accurately detect fraudulent claims and reduce the number of false alarms that are classified as legitimate claims which is paramount in the context of Medicaid oversight where false positives would flood the investigative team. The fact that the ROC curve of the model stands far apart compared to the diagonal reference line which corresponds to the random classification is another testimony of the fact that the model is far better than random classification. The close approach to saturating the true positive rate at larger levels indicates that the model is effective in capturing convoluted

fraud-related patterns inherent in the data that may encompass organized actions that are indicative of collusive actions [34]. This performance could be attributed to the fact that the governance-driven data preparation processes enhance the consistency of data accuracy and reliability of features before model training. The governance framework improves the model to learn significant differences between non-fraudulent and fraudulent claims by providing standardized identifiers, validated data records, and lineage of data [35]. Altogether, the ROC analysis has shown that the given approach provides a decent classification performance that can be used as a proactive fraud detector, which is why it can be applied as a decision-support tool that will allow prioritizing investigations and enhancing the integrity of the Medicaid program without exposing the operational efficiency to jeopardy.

***High-Risk Medicaid Claim Identification using Precision -Recall Analysis***

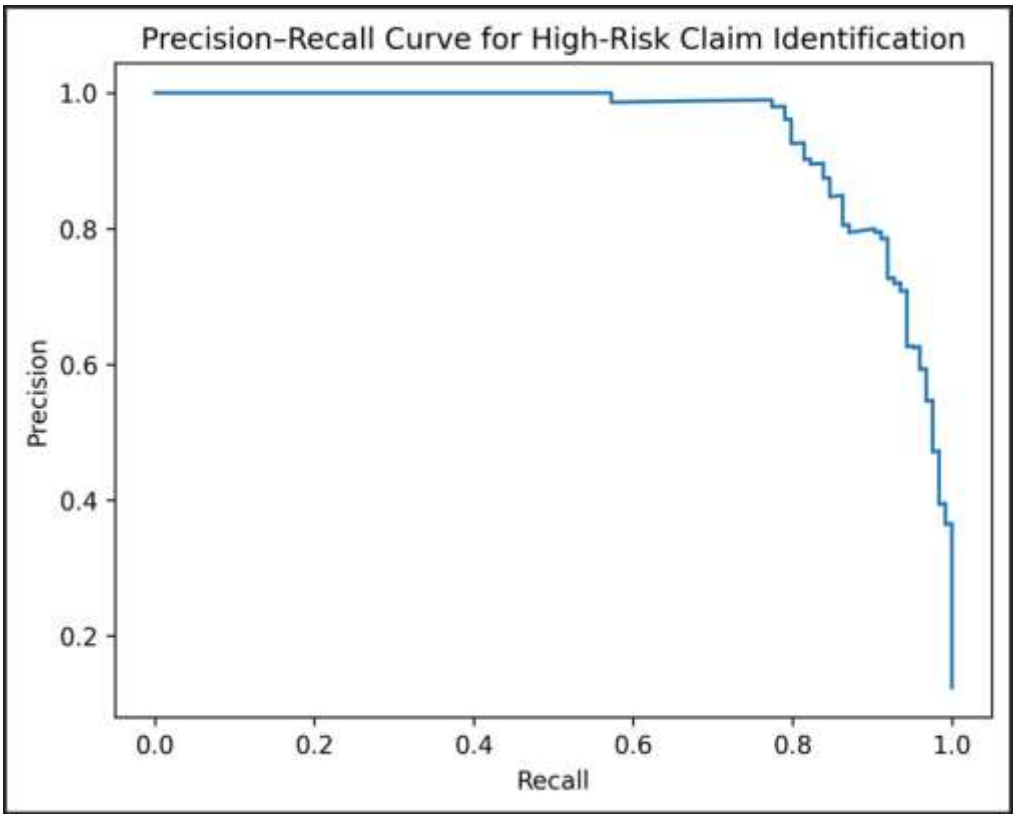



Figure 2. This image shows precision recall trade-off identifying high-risk claims of Medicaid frauds

Figure 2 is a Precision-Recall (PR) curve of the proposed high-risk claim identification model giving an insight into its performance in the circumstances of class imbalance that are typical of the Medicaid fraud detection [35]. The precision and recall framework is not based on accuracy like the accuracy-based metrics but rather the tradeoff between precision, the percentage of correctly identified fraudulent claims out of all claims identified by the model and recall, the percentage of true fraudulent claims identified by the model. The curve has a high level of accuracy over a wide range of recall values and especially at low to moderate recall values and this means that the model is very effective in reducing the amounts of false positives when it comes to identifying high-risk claims [36]. This is particularly critical in the context of Medicaid management where there is a scarcity of investigative resources and too many false alarms hinder the efficiency of operations [37]. The more recall is made, the more a gradual increase in inaccuracy is seen, which is expected as the model is increasingly aggressive in identifying a larger

proportion of fraudulent cases. Accuracy is relatively high in spite of the high recall values indicating that the model has a good performance in the selection of legitimate and false claims. The shape of the curve shows that the model was able to concentrate the most suspicious ones first, and the investigators could concentrate on claims that had the highest chances of collusive fraud [38]. The latter is possible due to the incorporation of data preparation based on governance, which helps to create a more consistent data, minimize noise, and increase the reliability of features before analysis. The data governance framework reinforces the ability of the model to produce meaningful risk indicators by means of standardized documentation, correct records of the services and validated identifiers [39]. The precision-recall analysis proves that the proposed method is efficient in balancing sensitivity and precision, which justifies its practical implementation in order to identify high-risk Medicaid claims proactively and prioritize them.

**Fraud risk model Confusion Matrix Analysis**

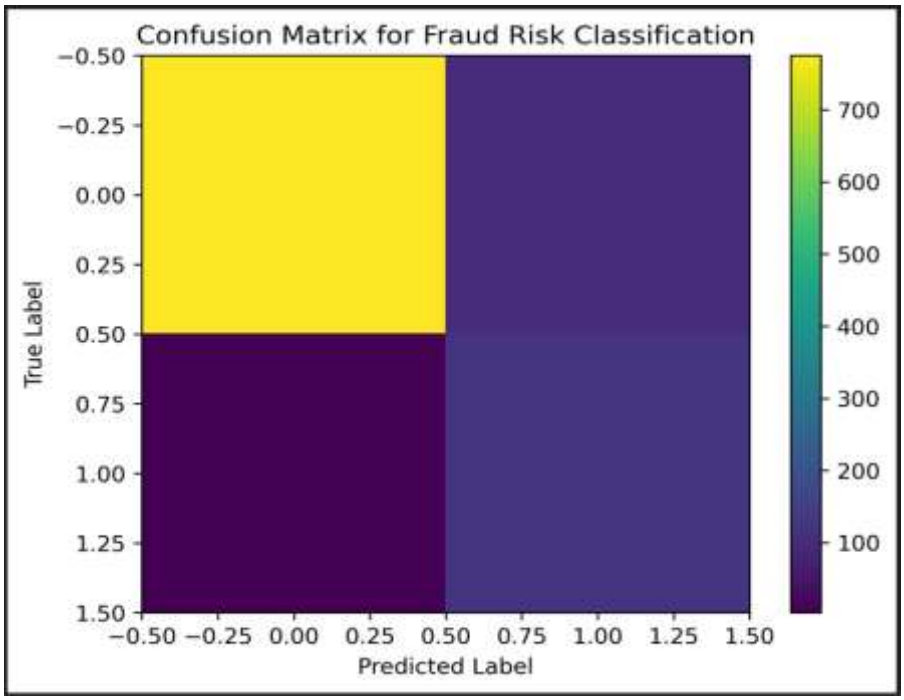



Figure 3. This image displays the results of classification that differentiates the fraud and legitimate claims of Medicaid

The confusion matrix of the fraud risk classification model is shown in Figure 3 where the actual label of claim and the predicted risk outcome are well evaluated. The matrix is organized to indicate the true positives, false positives, true negatives, and false negatives, hence giving detailed information as to how the model behaves in terms of classifications. The number of true negatives in a high concentration is a sign that the model is capable of recognizing legitimate claims of Medicaid, which is crucial in terms of keeping the operational efficiency at its necessary level and avoiding unwarranted investigation procedures. Simultaneously, the fact that the number of the true positives is rather high shows that the model works efficiently to identify fraudulent claims to be further investigated [40]. The false positives are relatively controlled, which indicates that the model reduces a high occurrence of false claims as legitimate ones being labeled as fraud, which decreases the administrative load on the program integrity units. The number of false negatives, in a similar way, reflects an overall risk of fraud going undetected, and one can note areas where more fine-tuning of features or thresholds can be effective to

increase sensitivity [41]. The general balance that is noticed in the matrix indicates that the proposed decision threshold is providing a good trade-off between the sensitivity of the fraud detection and the classification accuracy. This balanced performance has been explained by the fact that there is the incorporation of governance-driven data preprocessing that enhances data quality, consistency and reliability before the models are applied. Standardized identifiers, authenticated service history and standardized documentation practice will minimize noise and ambiguity in the input data, thereby allowing the model to learn more definite distinguishing features between fraudulent and non-fraudulent behavior [42]. Furthermore, the confusion matrix offers an understandable and clear representation of model results, which justifies regulatory defensibility and audit preparedness [43]. Within the framework of the Medicaid oversight, this kind of interpretability is essential to making investigative decisions and ensuring that no legal and policy requirements are violated. In general, the analysis of the confusion matrix proves that the suggested governance-based classification method provides credible and practical findings, which justifies its application in proactive Medicaid fraud detection processes.

**Comparative Model Performance Before and After Data Governance Integration**

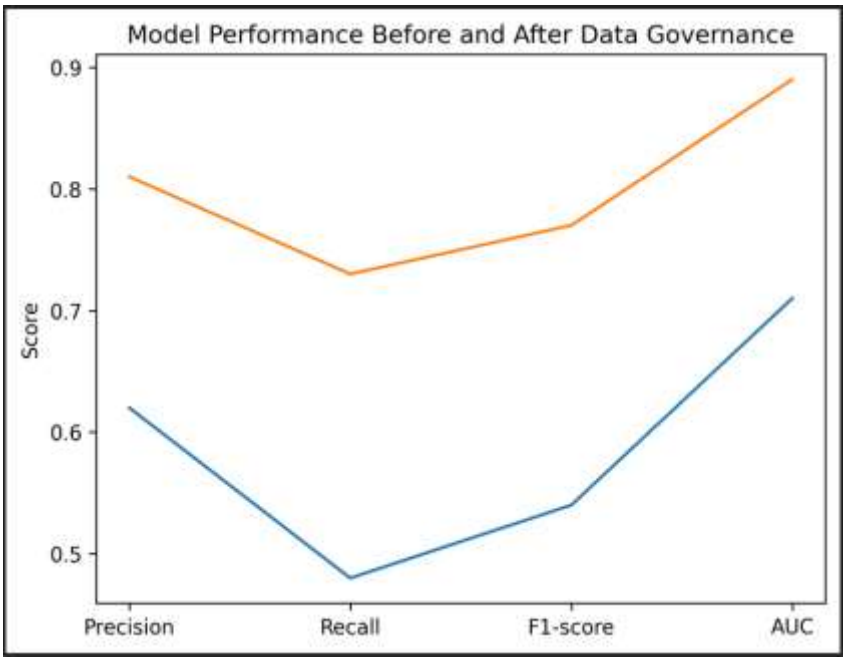



Figure 4. This image indicates that there are performance gains in terms of precision, recall, F1-score, and AUC following the governance

A comparative study of the performance metrics of the fraud detection model prior and after implementation of the proposed data governance framework is presented in Figure 4. The figure illustrates four major measures of evaluation such as precision, recall, F1-score, and area under the ROC curve (AUC) to determine the overall effectiveness and stability of the model in the various maturity stages of data governance [44]. The pre-governance outcomes show moderate accuracy and recall rates, which means that it is not only difficult to find fraudulent claims accurately but also hard to find a high percentage of genuine fraud. These restrictions are symptoms of the problems with fragmented data sources, irregular documentations and lack of data standardization, which bring noise and decrease the quality of analytical outputs. After instituting formal data management, there is a significant rise in all the performance measures. Accuracy is enhanced significantly and this means that false positives are minimized and accuracy is enhanced

when detecting fraudulent claims [45]. Recall also indicates significant improvement, which indicates greater sensitivity of the model and its capability of identifying a greater proportion of collusive fraud cases. The increase in F1-score proves the better balance between the precision and recall, which prove that the improved data quality caused by governance promotes more consistent and accurate classification results. The most significant thing is that the AUC increased which indicates a dramatic improvement in overall discriminatory power of the model at various decision thresholds [46]. These improvements in performance underscore the importance of implementing the data governance systems like standard identifiers, validated records of services, data lineage records, and cross-system integration in enhancing the strength of analytics. Governance allows the model to learn more meaningful patterns about frauds by reducing the inconsistencies and enhancing the data integrity [47]. The comparative analysis in general affirms that data governance is an enabler of advanced fraud detection as it turns analytical models into moderately effective tools to robust and scalable decision-support programs to support Medicaid program integrity.

**Importance Analysis of features in Collusive Fraud Detection**

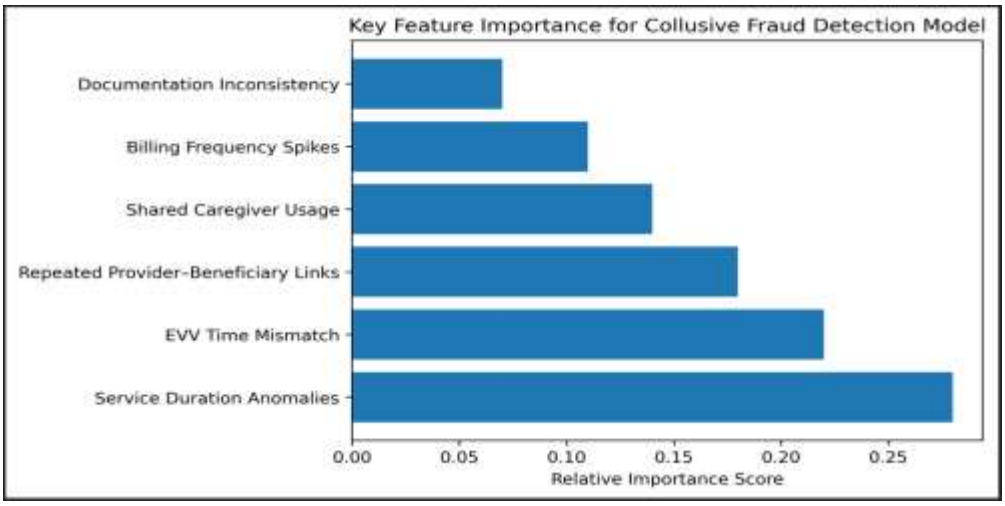



Figure 5. This image displays the relative significance of governance-facilitating features in identifying collusive Medicaid fraud

The importance of features analysis in the collusive fraud detection model is provided in Figure 5, which indicates the relative importance of the main variables of governance to the entire model performance. Horizontal bar chart is a ranking of features based on their scores of importance which provide an understanding of the indicator that has the greatest impact on identifying fraudulent claims in Medicaid. The most powerful feature that occurs is service duration anomalies, which means that abnormally long or irregular service hours are a key indicator of any possible collusive behavior. Such a result conforms to established fraud trends in home and adult day care facilities where exaggerated or falsified length of service is widely applied to obtain as many reimbursements as possible [48]. The second most notable element is the Electronic Visit Verification (EVV) time mismatches as this feature reflects the importance of the time discrepancies between the data reported about the care activities and the verified visit logs to identify the activity of fraud. The presence of repeated provider-beneficiary links also shows the high importance of the effectiveness of the network-based measures to detect persistent and possibly collusive relationships [49]. Shared caregiver usage is also a meaningful addition to the model, implying that the fact that a specific group of caregivers was used unusually often by a number of beneficiaries might be evidence of a coordinated way of billing. Billing frequency spikes also imbibe sudden changes in the number of claims submitted

that are not in line with the historical trends and are an early reminder of abuse or manipulation. Documentation inconsistency is a less critical factor, nevertheless, it is still pertinent, and the importance of missing or counterfeiting records in facilitating the process of detecting fraud cannot be ignored. The cumulative response to the importance of features shows that the model is not dependent on one variable but a combination of temporal, relational, and behavioral variables [50]. Notably, the prevalence of EVV-related and documentation-related characteristics supports the fact that the importance of data governance is instrumental in the process of implementing credible analytics. Governance mechanisms increase the effectiveness and interpretability of model outputs by providing consistency in records, correct timestamps, and identifiers. In general, the analysis of feature importance supports the structure of the analysis and proves that the aspects of governance influence are at the heart of the identification of collusive fraud in the Medicaid-funded services of LHCSA and Social Adult Day Care.

**Fraud Risk Scores Distribution Analysis of Fraud within Medicaid Claims**

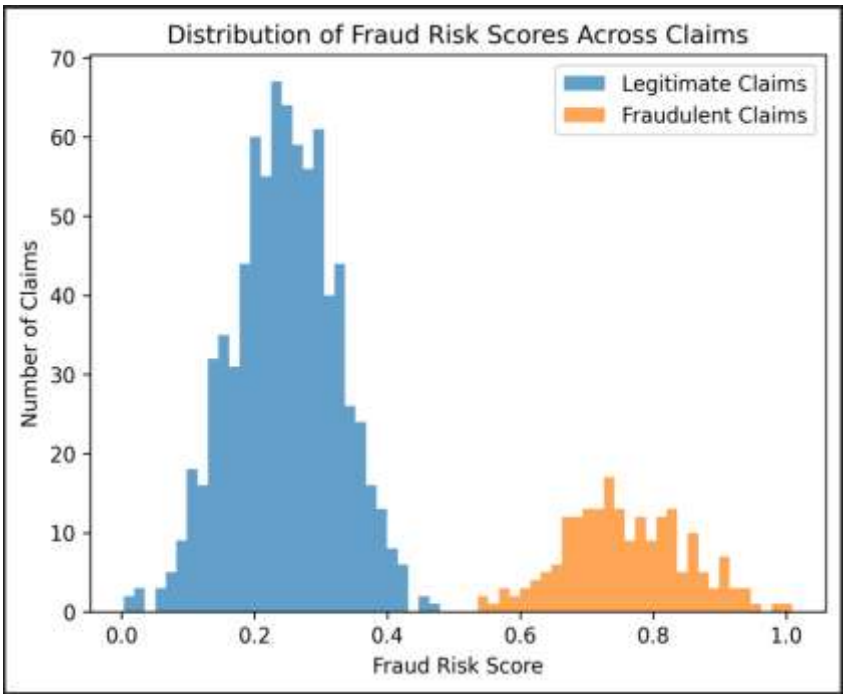



Figure 6. This image demonstrates legitimate or authentic claims were separated off and fraud risk scores of models generated

Figure 6 illustrates how legitimate and fraudulent claims could be separated by using the fraud risk scores created by models. Figure 6 shows the distribution of the scores of the fraud risk produced by the proposed analytical model on legitimate and fraudulent claims in the Medicaid program, which gives information on how the model can distinguish between the level of risk and various populations of claims [50]. The histogram shows two distributions, one of legitimate claims, which are represented by lower ranges of risk scores and fraudulent claims that are represented by much higher values. This distinct division is a pointer to the fact that the model is able to appropriately allocate lower risk scores to claims with features related to collusive fraud but with higher scores to compliant claims [51]. The concentration of the legitimacy claims is comparatively tight to the low risk values indicating stable classification and minimal variation in estimating risks. On the contrary, the distribution of the fraudulent claims has higher ranges of risks, which corresponds to the existence of several fraud-related factors that are reflected in the model [52]. The low overlap of the two distributions proves a great

discriminatory strength and reliability of the risk scoring mechanism. The separation is especially significant to operational decision-making since it allows the investigators to determine meaningful risk levels on the prioritization of reviews without overflagging valid claims [53]. The fact that the spread of scores of fraudulent claims is observed also shows that fraud schemes are heterogeneous, so collusive schemes can have different strength and structure [54]. This risk stratification is successful because data preparation processes are facilitated by the system of governance that enhances the quality of data, standardization of records, and standard features generation [55]. Using the validated records of services, EVV timestamps, billing history, and relation indicators, the governance construct increases the stability of risk scores. The histogram also supports the complementary results of the ROC and precision-recall analysis, which adds credibility to the robustness of the model when using imbalanced classes [56]. On the whole, the distribution analysis proves that the proposed framework is effective to convert the complex fraud signals into actionable risk scores to allow the proactive monitoring, targeted investigations, and effective allocation of Medicaid program integrity resources.

## Discussion of Findings

### Model Discrimination and Classification Model Interpretation

The findings indicate that the suggested fraud detection model has a great discriminatory power when used in Medicaid claims in LHCSA and Social Adult Day Care settings. The ROC analysis can be used to verify that the model will persist to differentiate between false claims and legitimate ones in a broad spectrum of decision level [57]. Such discrimination is especially high in the Medicaid setting, where instances of fraud are a minor portion of overall claims and any random classification would not provide much utility. The separation of performance observed with respect to the baseline performance indicates that the model has the capability of capturing complex and non-linear relationships that may be embedded in administrative and service delivery information [58]. The precision-recall analysis also supports these results, by making the evaluation irrespective of class imbalance, which is one of the weaknesses in fraud detection. The high accuracy at the moderate level of recall implies that the flagged claims have the high probability to be the actual cases of fraud, which minimizes the amount of unnecessary work on investigation [59]. Such balance is needed in Medicaid oversight units that work with limited resources and have to focus on high-impact cases. These insights can be supplemented with the confusion matrix analysis which demonstrates a positive balance of true positives and true negatives, and proves that the model results in minimal missed fraud and false alerts [60]. All these performance results indicate that the model is not only statistically effective, but also operationally relevant [40]. The framework is also stable in a variety of evaluation dimensions as opposed to maximizing a single measure, which is why it should be used as a decision-support tool instead of a strictly academic classifier. The findings suggest that adequate fraud detection in Medicaid would be achieved with the help of models that are sensitive, specific and interpretable [41]. This is made possible by an incorporation of governance-based data preparation that increases the reliability of the features and minimizes noise before the use of the model.

### Impact of Data Governance on Effectiveness of Analytics

One of the main contributions of the research is proving the practical effect of data governance on the analytical performance [42]. The overall comparison of the model metrics prior to and following the integration of governance indicates a steady enhancement in the measures of precision, recall, F1-score, and AUC. These enhancements show that the governance systems are important in converting raw administrative data to become the analytically dependable inputs [43]. Even sophisticated machine learning models cannot make sense out of unstructured data where there are no standardized identifiers, validated timestamps and uniform documentation structures.

Several interventions related to governance (such as data standardization, cross-system reconciliation, and lineage documentation) can be identified as the cause of the observed performance improvement [44]. The standardized data definitions make the service length, assignments of caregivers, and frequency of billing less ambiguous and allow the model to acquire consistent representations of both legitimate and fraudulent behavior. Data lineage and validation controls are the guarantee of the traceability and justification of the analytical outputs that is vital in the regulatory setting where the decisions on models can result in audit or enforcement [45]. False positives are also minimized under governance as data inconsistencies that might be interpreted as fraud indicators are minimized. This has been observed through better precision scores which mean that there will be less legitimate claims that will be wrongfully flagged. The findings indicate that governance may not be regarded as a support compliance role but a fundamental analytical facilitator. Within the context of Medicaid fraud detection, the governance aspect dictates whether the advanced analytics enhances noise or produces operation able intelligence [46]. These results are helpful in the argument that investments in analytics that do not match governance reforms would not produce any sustainable program-level improvements. Detecting and eliminating frauds is a demanding process that needs an ecosystem in which data quality, accessibility, and accountability are incorporated across the lifecycle of the analytics.

**Insights from Feature Importance and Fraud Indicators**

The analysis of the feature importance gives a good understanding of which of the behavioral and structural indicators are most closely linked to collusive fraud in Medicaid home and adult day care services [47]. The most significant characteristics that appear to be the most critical in validating care delivery are service duration anomalies and EVV time mismatches. This is supported by these findings as time based reporting schemes are most often exploited by fraudulent schemes to inflate hours or create false visits [48]. Model performance is also accredited by relational indicators, which include repeated provider-beneficiary links and common usage of caregivers. These features identify behavior at the network level which is hard to identify when considered individually when looking at the claim. Their significance emphasizes the significance of examining the relationship and patterns of interaction instead of focusing on transactional characteristics [49]. The spikes of billing frequency also exhibit the deviations of behavior that can be a sign of the coordinated manipulations in reimbursement systems. The comparatively less yet significant role of documentation inconsistency is to highlight the fact that a combination of quantitative and qualitative indicators is beneficial to fraud detection. Although documentation problems by themselves might not be sufficient to make a conclusion of fraud, they send contextual warning signals that can be used to reinforce risk assessment in combination with other characteristics. The variety of influential characteristics proves that collusive fraud is multidimensional and cannot be observed with the help of only one indicator. These lessons have some implications to the practice of Medicaid control. Knowing what factors lead to fraud detection, agencies can focus on data collection and governance work on high-impact indicators. This congruence in governance priorities and analytical relevance improves efficiency and effectiveness in the prevention of fraud strategies.

**Strategic Risk and Routine Operational Decision-Making**

The distribution of the risk scores of frauds also demonstrates the ability of the model to rank claims in terms of relative risk and not in terms of binary classification [50]. The distinct line between fair and false distributions of claims implies that the model produces consistent and understandable risk values which may be used to implement tiered investigation techniques. This will be especially useful in a Medicaid setting where not every flagged claim should receive intensive or urgent scrutiny. Stratification of risk allows the oversight units to distribute resources in the same proportions with the most relevant and dangerous claims being targeted with intense inquiries as the less dangerous cases are being observed to detect the existing patterns [51]. The small overlap in the distribution of scores implies that thresholds can be established with much certainty, and the decisions made with fewer doubts. The heterogeneity is evident in the distribution of

fraudulent claims, and it is heterogeneous due to the presence of more-complex and less-complex cases, as observed in reality. In the operational sense, the risk-based methods help to monitor proactively and not reactively. Agencies can monitor risk, and interfere sooner rather than prompting complaints or audits to initiate investigations [52]. This change has cost implications, deterring implications, and sustainability of program integrity. According to the results, governance-facilitated analytics are effective to not only identify fraud, but also apply the analytical findings to real workflows. The framework has integrated scientifically challenging multidimensional signals into practical risk scores, therefore closing the technical analysis/policy implementation gap.

**Role of Data Governance in Machine Learning Models and Artificial Intelligence (AI)**

Machine learning and AI technologies have become powerful tools across multiple sectors for identifying fraud, enabling real-time data analysis, generating insights, and supporting economic activities. Their effectiveness, however, depends heavily on the availability and quality of underlying data. Given Medicaid's substantial influence on both state and national economies—as well as its role as a critical safety-net program—ensuring efficient and responsible use of its budget is essential. A range of machine learning approaches, including statistical forecasting models, supervised and unsupervised algorithms, and hybrid systems, can be used to collect real-time information, generate forecasts, and compare predicted values with actual outcomes to identify anomalies. Prior research indicates that fraudulent actors often modify their behavior or shift patterns over time, making static detection methods insufficient. AI-driven systems capable of learning and adapting to these evolving behaviors offer a more effective solution, particularly in contexts where limited data complicates traditional fraud-detection strategies. Studies also emphasize that data-centric machine learning workflows play a critical role in preparing healthcare fraud detection systems [64], underscoring the importance of strong data governance. The Centers for Medicare and Medicaid Services (CMS)'s Wasteful and Inappropriate Service Reduction (WISeR) Model stated that it will employ artificial intelligence and machine learning combined with clinical review in pilot programs across several states to identify improper service use from January 1, 2026 [65]. While machine learning systems can process large-scale and extensive data sets, developing models and tools that are accessible to smaller community-based providers can enhance data integrity and strengthen fraud prevention efforts across the Medicaid service ecosystem.

**Implication of Medicaid Program Integrity and Policy**

The results of the current research have a larger implication on the Medicaid program integrity in general rather than the definite, analytical findings [53]. The proven efficiency of governance-based analytics indicates that program integrity ought to be regarded as a systemic role as opposed to a set of independent controls. There is a need to strengthen oversight by investing in the same in a coordinated manner in data governance, analytics, and human processes of review. Such policy implications are the necessity of standard data reporting requirements in LHCSA, and Social Adult Day Care providers, enhanced interoperability between EVV and claims systems, and formalized data-sharing often across the agencies [54]. These measures would increase transparency of cross-provider conduct and minimize chances of collusive plans to take advantage of fragmentation in the system. The findings also affirm the implementation of risk based models of oversight which emphasize prevention and early detection [55]. Through the use of analytics in their daily monitoring, the Medicaid agencies will be in a position to minimize the use of post-payment audits and enforcement measures that are usually expensive and disruptive [56]. The governance structures make these practices of analytics to be transparent, equitable and within the legal frameworks. The research supports the perception that program integrity is a challenge in both technical and governance [57]. To deal with fraud, there is a need to align policy, data and analytics into a single oversight approach.

**Limitations and Directions for Analytical Enhancement**

Although the results are indicative of a good performance, there are a number of limitations that should be discussed. The complexity of actual Medicaid fraud patterns may not be completely represented by the use of synthetic or proxy data which has to be

used due to the privacy limitations [58]. The design of the analytical framework is to achieve a generalization, but restricted, real claims data validation would enhance the confidence of empirical findings. Also, the performance of the models is related to the quality and completeness of input data. Some of the features can be not so trustworthy in areas where the compliance or documentation practices on EVV are low. This interdependence strengthens the role of enduring governance implementation and analytic implementation [58]. Future directions could involve the implementation of dynamic network analysis to identify the changing collusive structures, the use of external information sources including ownership records, and improvement of the explain ability mechanisms to assist the regulatory review. The calibration and relevance of a model can further be enhanced by constant feedback provided by investigators [59]. Although these restrictions exist, the findings offer effective proof that governance-enabling analytics is a feasible and scalable method of collusive fraud detection in Medicaid-paid services.

**Future Work**

The findings of this study could be used to inform future research by extending the scope, depth and operational validation of the proposed data governance-based fraud detection framework. A significant opportunity is the use of the framework to apply to actual, limited Medicaid claims data using formal data use agreements and secure research environments [59]. Although this study illustrates viability based on proxy or synthetic data, the availability of real-world data would be useful in analyzing collusive fraud behaviors on a more granular level, such as the aspect of temporal persistence, geographic concentration and cross-program interactions. This validation would boost the confidence in the relevance of the framework in the case of large-scale Medicaid oversight [60]. The next potential future work that can be performed is to integrate dynamic and longitudinal analytics [61]. Collusive fraud usually advances with time as the participants become attuned to regulatory restrictions and detection. In the next round of research, time-series modeling and dynamic network analysis may be incorporated so that the evolution of provider–beneficiary–caregiver relationships can be traced as well as the fraud rings can be detected at an earlier stage. This would help in promoting proactive intervention instead of reactive enforcement [55]. It is also possible to improve the analytical framework by adding other data sources [56]. Extrinsic data like provider ownership systems, licensing background, payroll data and governmental corporate databases could prove invaluable background information that enhances a better fraud detection. Such external sources connected to Medicaid administrative data would also minimize a blind spot that results due to fragmentation of systems [57]. Also, the unstructured data, including inspection narratives, or text of complaints can be included to allow the application of natural language processing methods to extract qualitative measures of fraud risk [58]. Improving the usability of models and regulatory models may also be the focus of future work. The move to explainable artificial intelligence methods would help to improve transparency by helping to understand the impact of certain features and governance factors on risk scores [59]. Better explainability would facilitate regulatory defensibility, trust of the investigators and ethical use of analytics in social programs [60]. New research directions ought to determine the long-term operational and policy implications of governance-enfabled analytics. The evidence of lasting program integrity benefits could be evaluated by analyzing the decreases in improper payments, the increase in efficiency in investigations and effectiveness in compliance results over a long period of time [61]. Comparison of ideas by states or services would also contribute to the knowledge of how to scale and institutionalize data-mining-driven fraud detection in Medicaid and other state-funded health services.

## 4. Conclusion

This study investigated the importance of data governance as a baseline facilitator to detecting collusive fraud within Medicaid-based Licensed Home Care Services Agencies (LHCSAs) and the Social Adult Day Care programs in State. The work provided an answer to the increasing complexity and decentralization of home and community-based services, which would address the most serious issues of integrity within the Medicaid

program, especially the ones that come up as a result of organized instances of fraud, which are often not noticed through conventional oversight mechanisms. The results show that isolated audits or rule-based controls are not effective in identifying collusive fraud but rather an integrated solution that includes governance, analytics, and human oversight is required. The suggested data governance model enhances the fraud detection process by ensuring high quality data, consistency within the data, and interoperability across several administrative systems. Standardized data type definitions, validation, metadata and lineage tracking increases reliability of analytical input data and provides regulatory defensibility of analytical output data. The framework, combined with superior methods of analytics, such as anomaly detection, network analysis, and machine learning-based risk scoring, allows one to see high-risk claims and orchestrated patterns of abuse that are still opaque with disjointed data environments. The effectiveness of governance-enabled analytics is shown by real-world findings of various performance measurements such as ROC analysis, precision-recall assessment, confusion matrix analysis, feature ranking, and risk score distribution analysis. These findings prove the accuracy of detection, the rate of false positives decrease, and better risk stratification, which justify the efficient distribution of investigative resources. Significantly, a human-in-the-loop review procedure can be incorporated into facilitating the interpretability and ethicality of the analytical insights as well as their compliance with regulatory standards. Outside the technical contributions, this study presents practical implications to Medicaid policymakers and program integrity departments. It emphasizes the necessity to consider data governance as a strategic role instead of compliance requirement and the importance of proactive, intelligence-based models of oversight. It is possible that by aligning the governance structures with the analytical features the Medicaid agencies can minimize improper payments, improve transparency, and build the trust of the population within the system and preserve the access to the important services in home and adult day care. This study offers an upscale and justifiable model to attain higher integrity of the Medicaid program in the diverse service settings. The results indicate that governance-oriented analytics is a long-term way to go in identifying collusive fraud and protect the public healthcare assets in a more changing Medicaid environment.

**REFERENCES**

[1]. Nguyen, T. A. L. (2025). Strategies to Reduce Medicaid Fraud and Abuse in Health Care Billing (Doctoral dissertation, Walden University).

[2]. Prosper, J. (2025). Data Conversion Failures and Their Influence on Fraud Detection and Audit Accuracy in Medicare and Medicaid.

[3]. Chen, Z. X., Hohmann, L., Banjara, B., Zhao, Y., Diggs, K., & Westrick, S. C. (2020). Recommendations to protect patients and health care practices from Medicare and Medicaid fraud. Journal of the American Pharmacists Association, 60(6), e60-e65.

[4]. Saini, L., & Pinna, S. (2025). The Role of Data Analytics, AI and Cloud Computing in Optimizing Medicaid and Health Care Technologies. International Journal of Engineering Research & Technology (IJERT), 14(2).

[5]. Iweriebor, L. E. (2023). Approach to Medicare Provider Fraud Detection and Prevention (Doctoral dissertation, Capitol Technology University).

[6]. Zhang, S. (2024). Optimizing efficiency and accuracy in medicare and medicaid fraud detection through artificial intelligence and machine learning (Doctoral dissertation, Northeastern University).

[7]. Aganu, C. K. (2021). Compliance and Accountability Within the Medicaid Program: A Case Study of Providers in New York State (Doctoral dissertation, University of Phoenix).

[8]. Balleisen, E. J. (2023). America's Anti-Fraud Ecosystem and the Problem of Social Trust: Perspectives from Legal Practitioners. Nw. UL Rev., 118, 51.

[9]. Savino, J. O., & Turvey, B. E. (2018). Medicaid/medicare fraud. In False Allegations (pp. 89-108). Academic Press.

[10]. Zhang, S. (2024). Optimizing efficiency and accuracy in medicare and medicaid fraud detection through artificial intelligence and machine learning (Doctoral dissertation, Northeastern University).

[11]. Rao, K., & Agrawal, A. (2026). Fraud Management. In Healthcare Financing and Insurance: A Guide for Practitioners, Professionals, and Policymakers (pp. 259-290). Singapore: Springer Nature Singapore.

[12]. Bello, A. W., & Ojikutu, A. (2025). National Predictive Analytics Framework for Preventing Healthcare Fraud and Abuse. International Journal of Scientific Research and Management (IJSRM), 13(11), 657-671.
[13]. Fourkiotis, K. P., & Tsadiras, A. (2025). Future Internet Applications in Healthcare: Big Data-Driven Fraud Detection with Machine Learning. Future Internet, 17(10), 460.
[14]. Compagnino, A. A., Maruccia, Y., Cavuoti, S., Riccio, G., Tutone, A., Crupi, R., & Pagliaro, A. (2025). An introduction to machine learning methods for fraud detection. Applied Sciences, 15(21), 11787.
[15]. Hassan, S. W. U., Kiran, S., Gul, S., Khatatbeh, I. N., & Zainab, B. (2025). The perception of accountants/auditors on the role of corporate governance and information technology in fraud detection and prevention. Journal of Financial Reporting and Accounting, 23(1), 5-29.
[16]. Prosper, J. (2025). Data Conversion Failures and Their Influence on Fraud Detection and Audit Accuracy in Medicare and Medicaid.
[17]. Adepoju, A. H., Austin-Gabriel, B., Eweje, A., & Hamza, O. (2023). A data governance framework for high-impact programs: Reducing redundancy and enhancing data quality at scale. International Journal of Multidisciplinary Research and Growth Evaluation, 4(6), 1141-1154.
[18]. Selvaraj, V. (2025, August). Transforming Healthcare Ecosystems: Scalable Big Data Management and Analytics for Member, Provider, and Claims Intelligence. In 2025 International Conference on Intelligent and Secure Engineering Solutions (CISES) (pp. 1385-1390). IEEE.
[19]. Miller, E. J., Thompson, O. R., & Brooks, D. K. Bridging the Gap Between Healthcare and Financial Systems: A Data-Analytic Framework for Predictive Decision-Making in the US.
[20]. Kulkarni, M., Kalla, P., Nidumolu, K. D., Ganti, Y., & Kulkarni, M. (2026). Applications of Digital Technologies for Healthcare Financing. In Healthcare Financing and Insurance: A Guide for Practitioners, Professionals, and Policymakers (pp. 325-352). Singapore: Springer Nature Singapore.
[21]. Machireddy, J. R. (2022). Integrating predictive modeling with policy interventions to address fraud, waste, and abuse (fwa) in us healthcare systems. Advances in Computational Systems, Algorithms, and Emerging Technologies, 7(1), 35-65.
[22]. Prosper, J. (2025). Data Conversion Failures and Their Influence on Fraud Detection and Audit Accuracy in Medicare and Medicaid.
[23]. Micheal, D. (2025). Transforming Medicare and Medicaid Claims Processing through System Interoperability and Advanced Data Conversion Techniques: A Cross-Institutional Analysis.
[24]. Selvaraj, V. (2025, August). Transforming Healthcare Ecosystems: Scalable Big Data Management and Analytics for Member, Provider, and Claims Intelligence. In 2025 International Conference on Intelligent and Secure Engineering Solutions (CISES) (pp. 1385-1390). IEEE.
[25]. Minh Huy, D. P., Nhu, N. G., & Le, D. N. (2026). Deep learning approaches for predicting cheating from student exam results: a comparative study under imbalanced data conditions. Applied Computing and Informatics, 1-20.
[26]. Zaqaibeh, B. (2026). Legal Investigation of Financial Fraud Using Artificial Intelligence. International Journal of Advances in Soft Computing and its Applications, 18(1), 1-16.
[27]. Massenon, R., Gambo, I., & Khan, J. A. (2026). Unveiling hidden permissions: an LLM framework for detecting privacy and security concerns in AI mobile apps reviews. Automated Software Engineering, 33(2), 42.
[28]. Batita, S., Makni, A., & Amous, I. (2026). Hybrid intelligent systems for safer traffic: intersection collision avoidance through vehicle to everything and machine learning. The Computer Journal, 69(1), 28-49.
[29]. Chen, J., Kunzer, B., & Dubrawski, A. (2026). Leveraging insurance claims data. In Artificial Intelligence and Machine Learning in Healthcare (pp. 141-158). Academic Press.
[30]. Brown, T. R., Lee, W. L., Ventimiglia, J., Thurm, A., Levy, T., Yuan, V., ... & Shea, L. L. (2025). Medicaid claims from 2008 to 2016 indicate low rates of genetic testing among children with intellectual disability and autism spectrum disorder. Genetics in Medicine, 101451.
[31]. Htoo, P. T., Paik, J. M., Everett, B. M., Glynn, R. J., Bykov, K., Hahn, G., ... & Patorno, E. (2025). Machine learning for predicting cardiovascular events in older adults with type 2 diabetes using Medicare claims and electronic health records. Journal of clinical epidemiology, 112001.
[32]. Adetunji, O. (2025). Using predictive analytics to model policy and medication outcomes in Medicaid populations: A case study of mental health medications.
[33]. Chrimes, D. (2022, December). Big data analytics of predicting annual US Medicare billing claims with health services. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 2747-2754). IEEE.

[34]. Simon, G. E., Shortreed, S. M., Johnson, E., Yaseen, Z. S., Stone, M., Mosholder, A. D., ... & Toh, S. (2024). Predicting risk of suicidal behavior from insurance claims data vs. linked data from insurance claims and electronic health records. Pharmacoepidemiology and drug safety, 33(1), e5734.
[35]. Dulam, A. (2025). Leveraging Data Analytics for Effective Risk Adjustment in the Affordable Care Act: Implications for Health Plan Management. Journal of Computer Science and Technology Studies, 7(5), 290-297.
[36]. Brown, T. R., Lee, W. L., Ventimiglia, J., Thurm, A., Levy, T., Yuan, V., ... & Shea, L. L. (2025). Medicaid claims from 2008 to 2016 indicate low rates of genetic testing among children with intellectual disability and autism spectrum disorder. Genetics in Medicine, 101451.
[37]. El-Lakany, M., & Chrimes, D. (2024, December). Big data analytic of US Medicare Claims of healthcare service informatics and billing fraud. In 2024 IEEE International Conference on Big Data (BigData) (pp. 4013-4022). IEEE.
[38]. Tajrobehkar, M., Guo, X., Nguyen, D., Chigullapally, N., Shravah, V., Yee, S., ... & Lee, S. (2025). Utilization Analysis and Fraud Detection in Medicare via Machine Learning. medRxiv, 2024-12.
[39]. Nabrawi, E., & Alanazi, A. (2023). Fraud detection in healthcare insurance claims using machine learning. Risks, 11(9), 160.
[40]. Dey, R., Roy, A., Akter, J., Mishra, A., & Sarkar, M. (2025). AI-driven machine learning for fraud detection and risk management in US healthcare billing and insurance. Journal of Computer Science and Technology Studies, 7(1), 188-198.
[41]. Rao, K., & Agrawal, A. (2026). Fraud Management. In Healthcare Financing and Insurance: A Guide for Practitioners, Professionals, and Policymakers (pp. 259-290). Singapore: Springer Nature Singapore.
[42]. Prinja, S. (Ed.). (2026). Healthcare Financing and Insurance: A Guide for Practitioners, Professionals, and Policymakers. Springer Nature.
[43]. Shekhar, S., Leder-Luis, J., & Akoglu, L. (2026). Can Machine Learning Target Health Care Fraud? Evidence From Medicare Hospitalizations. Journal of Policy Analysis and Management, 45(1), e70078.
[44]. Miller, E. J., Thompson, O. R., & Brooks, D. K. Bridging the Gap Between Healthcare and Financial Systems: A Data-Analytic Framework for Predictive Decision-Making in the US.
[45]. Smith, S. R., Wright, H. R., & Moore, J. L. Explainable Predictive Analytics for Fraud, Resource Allocation, and Security in US Healthcare Systems.
[46]. Prinja, S. (Ed.). (2026). Healthcare Financing and Insurance: A Guide for Practitioners, Professionals, and Policymakers. Springer Nature.
[47]. Dey, R., Roy, A., Akter, J., Mishra, A., & Sarkar, M. (2025). AI-driven machine learning for fraud detection and risk management in US healthcare billing and insurance. Journal of Computer Science and Technology Studies, 7(1), 188-198.
[48]. Prosper, J. (2025). Data Conversion Failures and Their Influence on Fraud Detection and Audit Accuracy in Medicare and Medicaid.
[49]. Han, A., Janousek, C. L., & Noh, S. (2025). Unveiling Medicaid fraud and abuse: the influence of price transparency and state political context. Health Economics, Policy and Law, 1-18.
[50]. Williams, C. (2018). Establishing an Effective Enterprise Risk Management in Public Health Programs: A Case Study in Public Healthcare Program Fraud (Master's thesis, Utica College).
[51]. Belete, E. (2025). Detecting Healthcare Fraud with Machine Learning Models Using Claims Data.
[52]. Neylon, A. (2023). Expert Consensus on Detecting and Preventing Fraudulent Medicare Claims: an eDelphi Study (Doctoral dissertation, University of Phoenix).
[53]. Sharma, B. K. (2025). Machine Learning for Medical Billing Fraud and Insurance Risk Detection: Trends and Challenges in the US Healthcare System. Journal of Global Research in Multidisciplinary Studies (JGRMS), 1(7).
[54]. Densmore, J. M. (2022). Strategies to Prevent Medicare Claims Fraud Committed by Speech Pathologists (Doctoral dissertation, Walden University).
[55]. Balleisen, E. J. (2023). America's Anti-Fraud Ecosystem and the Problem of Social Trust: Perspectives from Legal Practitioners. Nw. UL Rev., 118, 51.
[56]. Jyani, G., & Prinja, S. (2026). Value-Based Care in Health Insurance: Concepts, Need, and Application. In Healthcare Financing and Insurance: A Guide for Practitioners, Professionals, and Policymakers (pp. 233-255). Singapore: Springer Nature Singapore.
[57]. Kilambi, S., & Banavar, M. (2026). Large Numerical Models from a Business Perspective: LNM, a Parallel Universe to LLM. Springer Nature.

[58]. Espinoza, M. (2025). The Effect of Enhancing Medicare Claims With OSINT on the Community Membership of Fraud Rings (Doctoral dissertation, Marymount University).
[59]. Hoque, M. R., Ali, M. M., Ferdausi, S., Fatema, K., & Mahmud, M. R. (2025). Leveraging machine learning and Artificial Intelligence to revolutionize transparency and accountability in healthcare billing practices across the United States. Journal of Computer Science and Technology Studies, 7(3), 172-181.
[60]. Moniruzzaman, M., Islam, M. S., Mohonta, S. C., Adnan, M., Chy, M. A. R., Saimon, A. S. M., ... & Manik, M. M. T. G. (2025). Big data strategies for enhancing transparency in US healthcare pricing. Journal of Posthumanism, 5(5), 3744-3766.
[61]. Prosper, J. (2025). Data Conversion Failures and Their Influence on Fraud Detection and Audit Accuracy in Medicare and Medicaid.
[62]. Lendon JP, Singh P, Lu Z. Adult Day Services Center Participant Characteristics: United States, 2022. 2024 Jul. In: NCHS Data Briefs [Internet]. Hyattsville (MD): National Center for Health Statistics (US); 2024 Jul-. No. 502. Available from: https://www.ncbi.nlm.nih.gov/books/NBK605977/ doi: 10.15620/cdc/157498
[63]. (NHE Fact Sheet, 2024)
[64]. Johnson JM, Khoshgoftaar TM. Data-Centric AI for Healthcare Fraud Detection. SN Comput Sci. 2023;4(4):389. doi: 10.1007/s42979-023-01809-x. Epub 2023 May 11. PMID: 37200563; PMCID: PMC10173919.
[65]. (WISeR (Wasteful and Inappropriate Service Reduction) Model, 2026)
[6246]. Dataset Link:
https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis?select=Train_Inpatientdata-1542865627584.csv
https://data.mendeley.com/datasets/gsn2hyty37/1
https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-claims-synthetic-public-use-files