International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.2 (2025) pp. 3759-3771
http://www.ijcesen.com
ISSN: 2149-9144

Research Article

AI-Driven Eligibility Verification in U.S. Medicaid: A T-MSIS V4-Aligned Real-Time Data Matching Framework for Reducing Improper Payments and Protecting Public Program Integrity

Asif Hasan[1*], Abu Nasir[2], Nasrin Sultana[3]

[1]MS in Business Analytics Feliciano School of Business Montclair State University, NJ, USA
* **Corresponding Author Email:** asif.mdhasan92@gmail.com- **ORCID:** 0009-0003-3506-8910

[2]Master of Arts in Information Technology Management George Herbert Walker School of Business and Technology Webster University, St. Louis, Missouri, USA
Email: irfannasir000@gmail.com- **ORCID:** 0009-0008-9277-0454

[2]Master of Arts in Information Technology Management George Herbert Walker School of Business and Technology Webster University, St. Louis, Missouri, USA
Email: nsultaana94@gmail.com- **ORCID:** 0009-0008-6855-3732

Article Info:

DOI: 10.22399/ijcesen.5495
Received : 02 January 2025
Revised : 20 February 2025
Accepted : 25 February 2025

Keywords

Medicaid,
T-MSIS,
eligibility verification,
artificial intelligence,
record linkage,
program integrity

Abstract:

Medicaid experiences billions of dollars in annual improper payments arising from eligibility determination errors, inaccurate beneficiary information, and fraudulent enrollment. Fragmented state eligibility systems and sequential manual verification hinder timely detection. The CMS Transformed Medicaid Statistical Information System Version 4 (T-MSIS V4) provides standardized national data infrastructure, but current implementations support reporting rather than intelligent verification. This study proposes and empirically evaluates MedVerify-AI, an artificial intelligence-driven eligibility verification framework aligned with T-MSIS V4 that combines multi-source probabilistic record linkage with AI-based anomaly scoring for real-time detection of improper enrollments. The framework links applications to external verification sources (identity, income, death, residency records) using Fellegi–Sunter probabilistic record linkage, computes verification-layer scores across identity, financial, status, and duplicate dimensions, and fuses evidence with a supervised meta-learner. Evaluation used a reproducible synthetic dataset of 61,096 Medicaid applications over 24 months containing 5,902 anomalous applications (9.7%) across five documented anomaly types, with a temporal train/test split and four baselines: rule-based verification, logistic regression, isolation forest, and random forest. Probabilistic linkage achieved F1 of 0.998, recovering most identity links missed by deterministic matching (recall 0.996 vs. 0.988) at 1.18 ms per record. The proposed framework achieved the highest F1-score (0.803) and ROC-AUC (0.884) of all methods, improving recall by 13.5 percentage points over rule-based verification (0.727 vs. 0.593) at a 0.91% false-positive rate. Deceased-enrollee, income-misreporting, and residency anomalies were detected with complete recall. Combining standardized T-MSIS data structures, probabilistic linkage, and AI-based evidence fusion materially outperforms rule-based verification for pre-enrollment screening. Validation on operational Medicaid data is the essential next step.

1. Introduction

Medicaid is the largest public health insurance program in the United States, providing essential healthcare coverage to low-income families, older adults, individuals with disabilities, and other vulnerable populations. Jointly funded by the federal and state governments, Medicaid has expanded substantially over the past decade, resulting in increased enrollment, higher expenditures, and greater administrative complexity. While this growth has improved access to healthcare services, it has also intensified the challenge of ensuring that only eligible individuals receive benefits and that public funds are distributed appropriately (Centers for Medicare & Medicaid Services [CMS], 2025a; CMS, 2024a).

As Medicaid enrollment has increased, program integrity has become a major concern for federal and state agencies. According to the Fiscal Year 2024 Payment Integrity Dataset, Medicaid continues to account for one of the largest categories of improper payments across U.S. government programs, with billions of dollars attributed to eligibility errors, documentation deficiencies, administrative mistakes, and fraudulent activities (PaymentAccuracy.gov, 2024). Consequently, improving eligibility verification has become a strategic priority for CMS, the Office of Inspector General (OIG), and the U.S. Government Accountability Office (GAO) (GAO, 2024).

Oversight reports have identified persistent weaknesses in Medicaid data quality and verification. The OIG reported that states do not consistently define or validate paid amount data for Medicaid managed care drug claims (Office of Inspector General, 2024), and that multiple states continued making Medicaid capitation payments after beneficiaries had died — a direct failure of cross-agency beneficiary status verification (Office of Inspector General, 2023). Earlier GAO assessments concluded that CMS should strengthen documentation requirements and verification procedures to reduce improper payments (GAO, 2019, 2020).

T-MSIS V4 provides a standardized national framework for collecting, validating, and reporting Medicaid data across all participating states, establishing consistent data structures, standardized identifiers, eligibility variables, and reporting requirements (CMS, 2024a, 2025a). The accompanying T-MSIS Data Dictionary and technical guidance define standardized data elements, beneficiary identifiers, and eligibility records that create a common data infrastructure for nationwide Medicaid administration (CMS, 2024a, 2024b, 2024c, 2024d, 2025a). Despite these advances, eligibility verification remains largely dependent on fragmented workflows relying on manual document review and sequential verification across agencies such as the Internal Revenue Service and the Social Security Administration, delaying verification and reducing the ability to identify inconsistent or fraudulent applications before benefits are approved (CMS, 2024e; GAO, 2020).

Recent advances in artificial intelligence offer opportunities to modernize public-sector eligibility verification. AI-driven approaches have demonstrated potential for decision support, anomaly detection, and workflow automation across healthcare systems (Ingole et al., 2024); fairness-aware computational methods have been proposed for Medicaid eligibility determination (Fang et al., 2019); and integrated eligibility systems have been associated with improved participation and administrative coordination across public assistance programs (Mandal, 2025). A critical enabling capability is record linkage: the probabilistic framework of Fellegi and Sunter (1969), extended by Christen (2012) into practical methods for data matching and entity resolution, provides the methodological foundation for matching applicant records across heterogeneous federal and state repositories.

Although previous studies have investigated Medicaid program integrity and AI applications in healthcare administration, few have integrated standardized T-MSIS V4 data structures with real-time AI-driven verification and multi-source probabilistic matching in a unified, empirically evaluated framework. Existing approaches emphasize post-enrollment audits and retrospective recovery rather than proactive pre-enrollment screening. This paper addresses that gap. Its contributions are threefold: (1) a T-MSIS V4-aligned verification architecture combining probabilistic record linkage, multi-source verification layers, and AI-based evidence fusion; (2) a reproducible synthetic evaluation environment of 61,096 Medicaid applications with five injected anomaly types grounded in documented OIG and GAO findings; and (3) quantitative evidence that the framework achieves the best F1-score and ROC-AUC among all evaluated methods, improving detection recall by 13.5 percentage points over rule-based verification while maintaining a sub-1% false-positive rate.

1.1 Research Problem

Despite continuous investment in Medicaid modernization and data standardization, eligibility verification remains one of the most significant operational and financial challenges in the U.S. healthcare system. State programs rely on diverse information systems, varying data collection practices, and independent determination processes, creating barriers to timely verification and increasing the likelihood of eligibility errors, duplicate enrollments, and improper payments (CMS, 2025a; CMS, 2024a).

1.2 Research Gap

Government reports have consistently recommended strengthening eligibility verification and data quality (GAO, 2019, 2020, 2024; Office of

Inspector General, 2023, 2024), and CMS has advanced standardized data collection through T-MSIS V4 (CMS, 2024a, 2025a). What remains missing is an empirically evaluated framework that integrates this standardized infrastructure with AI-driven, real-time, multi-source verification — the gap this study addresses with a quantitative, baseline-controlled evaluation.

1.3 Research Objectives

The primary objective is to develop and empirically evaluate MedVerify-AI, an AI-driven eligibility verification framework aligned with T-MSIS V4. The specific objectives are to: (1) examine the challenges in existing Medicaid eligibility verification contributing to improper payments; (2) design a T-MSIS V4-compliant data normalization framework; (3) implement a multi-source matching engine verifying applicant information across federal and state sources; (4) integrate probabilistic record linkage and measure its accuracy against deterministic matching; (5) develop an AI-based anomaly scoring module for statistically unusual eligibility patterns; (6) evaluate the framework against rule-based and machine learning baselines using precision, recall, F1-score, ROC-AUC, false-positive rate, and processing efficiency; and (7) demonstrate how AI-supported verification can strengthen Medicaid program integrity.

2. Literature Review

2.1 Medicaid Eligibility Verification

Medicaid eligibility verification ensures public healthcare benefits go to individuals meeting federal and state requirements, checking applicant income, household size, citizenship or immigration status, residency, and related criteria. Because Medicaid is administered jointly, each state operates its own determination system under federal CMS rules, producing significant variation in verification methods, data quality, and system interoperability (CMS, 2024a; CMS, 2025a). Historically, determination relied on manual paperwork and caseworker review; even with electronic verification against federal and state databases, many systems depend on sequential workflows retrieving information from independent agencies, increasing administrative workload, delaying decisions, and obscuring inconsistencies until after benefits are approved (CMS, 2024b).

Federal oversight has repeatedly flagged these weaknesses. GAO found that documentation standards and verification procedures need strengthening to reduce improper payments (GAO, 2019), that eligibility determination errors continue to drive payment inaccuracies (GAO, 2020), and most recently that further actions are needed to enhance program integrity across Medicare and Medicaid (GAO, 2024). The OIG documented inconsistent state validation of managed care payment data (Office of Inspector General, 2024) and capitation payments continuing after beneficiaries' deaths due to flawed cross-agency status verification (Office of Inspector General, 2023) — the latter directly motivating the death-record verification layer evaluated in this study.

To promote national consistency, CMS launched T-MSIS V4, establishing standardized reporting requirements and data structures for state Medicaid agencies, supported by the T-MSIS Data Dictionary, technical instructions, provider location guidance, and eligibility-related definitions (CMS, 2024a, 2024b, 2024c, 2024d, 2025b, 2025c). However, T-MSIS V4 primarily supports standardized reporting rather than intelligent verification: current systems rely on fixed business rules and manual review, limiting real-time cross-agency verification. This creates the opportunity to use standardized T-MSIS V4 data as the foundation for AI-assisted verification supporting accurate, timely, and scalable enrollment decisions.

2.2 T-MSIS V4 Data Standard and Interoperability

T-MSIS V4 is the current national standard for Medicaid and CHIP data reporting. Its modular, file-oriented architecture receives standardized submissions from state Medicaid Management Information Systems, validates each against predefined business rules, and integrates them into the national CMS repository. Standardized file groups — beneficiary eligibility, provider information, managed care enrollment, claims, financial transactions, and third-party liability — are connected through common identifiers supporting longitudinal analysis. The data model specifies standardized definitions, controlled vocabularies, validation rules, and relational consistency, with metadata defining permissible values and reporting requirements for each element. These properties matter for intelligent verification in three ways. First, standardized reporting improves the data quality on which analytical models depend. Second, standardized beneficiary and provider identifiers make cross-dataset linkage tractable — essential for detecting duplicate enrollments and conflicting eligibility records. Third, the consistent structure provides the technological base advanced analytics require: AI, machine learning, and anomaly detection need consistent, high-quality data to produce reliable results. Interoperability, however, extends beyond format standardization to the ability of independent systems to share and interpret information; current T-MSIS V4 implementations do not perform automated eligibility verification, probabilistic identity matching, or AI-driven anomaly detection during enrollment. The MedVerify-AI framework evaluated here builds directly on this standardized foundation, adding the intelligent decision-support layer T-MSIS V4 currently lacks.

2.3 Record Linkage and AI Methods for Eligibility Verification

Accurately matching applicant records across independent databases is the methodological core of multi-source verification. The probabilistic record linkage framework of Fellegi and Sunter (1969) formalizes the decision of whether two records refer to the same individual as a likelihood-ratio test: for each comparison field, agreement or disagreement contributes a weight $\log_2(m/u)$, where m is the probability of agreement among true matches and u the probability of agreement among non-matches, and the summed weight is compared against acceptance thresholds. Christen (2012) extended this foundation into modern practice — blocking and indexing strategies that make pairwise comparison tractable at administrative scale, approximate string comparison functions tolerant of typographical variation, and evaluation methodology for linkage quality. These techniques are directly relevant to Medicaid because eligibility determination requires matching applicant information across federal and state repositories whose records differ in completeness, formatting conventions, and error rates: exact-key matching fails precisely where identities are recorded with clerical noise or deliberately manipulated, which is where verification matters most.

AI applications in Medicaid administration are an active research area. Ingole et al. (2024) survey AI-driven innovation across Medicaid access, cost efficiency, and population health management, identifying eligibility operations as a high-value automation target. Fang et al. (2019) propose fairness-aware computational methods for Medicaid eligibility determination, constructing fair groups to reduce bias while maintaining decision consistency — a consideration that becomes binding whenever machine learning informs public benefit decisions. Mandal (2025) provides causal evidence that integrated eligibility systems — shared verification infrastructure across Medicaid and SNAP — improve program participation and administrative coordination, supporting the premise that verification infrastructure investments yield returns beyond fraud reduction. What this literature lacks, and what the present study contributes, is an integrated, quantitatively evaluated pipeline connecting standardized T-MSIS data structures, probabilistic linkage, and supervised anomaly detection with head-to-head baseline comparison.

3. Proposed MedVerify-AI Framework

3.1 Framework Overview

MedVerify-AI comprises five connected phases — data acquisition and integration, preprocessing, record linkage and verification, AI-assisted risk assessment, and decision support — each compatible with the standardized data structures defined by T-MSIS V4. The first phase ingests beneficiary eligibility records structured to the T-MSIS V4 data model along with external administrative verification sources. Preprocessing then improves data quality through cleaning, normalization, and validation (Table 2 summarizes the pipeline). Record linkage compares applicant records across sources using deterministic and probabilistic matching, identifying duplicate identities, inconsistent demographic information, and conflicting eligibility records that isolated database checks miss. The AI-assisted risk assessment phase does not replace administrative decision-making; it scores records and prioritizes applications for specialist review while routine cases follow established workflows. The final phase presents verification results, confidence scores, and supporting evidence for transparent eligibility review. Figure 2 illustrates the workflow.

3.2 System Architecture

The architecture consists of six logical layers, illustrated in Figure 3. The Data Acquisition Layer receives beneficiary eligibility information in T-MSIS V4 format along with administrative information from external government verification systems. The Integration Layer unifies data into a T-MSIS V4-compliant format via standardized field mapping, schema validation, and quality checks. The Preprocessing Layer performs cleaning, normalization, duplicate detection, missing-value treatment, and attribute standardization. The Record Linkage Layer executes deterministic and probabilistic matching to find duplicate identities, conflicting demographics, and cross-database discrepancies. The AI Decision-Support Layer applies machine-learning risk scoring and anomaly detection to the linked records — focusing specialist attention rather than replacing administrative judgment. Finally, the Reporting and Decision Layer presents verification results, confidence scores, and evidence through a structured decision-support interface used by eligibility specialists to guide manual review and record final determinations.

4. Methodology

4.1 Research Design

This study employs a design-oriented research approach to develop and empirically evaluate MedVerify-AI. Because privacy regulations restrict access to operational Medicaid administrative data, evaluation uses a synthetic dataset constructed to reflect documented eligibility-verification failure modes from OIG and GAO oversight reports — including capitation payments after enrollees' deaths (Office of Inspector General, 2023), eligibility determination errors (GAO, 2020), and duplicate enrollment vulnerabilities (GAO, 2024). The simulation is deterministic given a fixed random seed, making the evaluation environment fully reproducible.

4.2 Synthetic Dataset Construction

The simulation generates 61,096 Medicaid enrollment applications over a 24-month period, each carrying applicant-supplied identity fields (name, date of birth, Social Security number), reported income, and documentation status. Legitimate applicant incomes follow a log-normal distribution truncated to plausible ranges, with 85% of applicants drawn below the simplified annual eligibility threshold ($20,120) and reported income varying from verified income by normal clerical error of approximately ±3%. Realistic recording noise is applied to application identities: character substitutions, transpositions, and truncations in names (12% of records), and day/month swaps, off-by-one days, and off-by-one years in dates of birth (8%). Five external verification sources are simulated with their own independent recording noise: a Social Security Administration identity master file (4–5% field noise), IRS income records, the Death Master File, state residency records, and prior enrollment history — so that no source is a perfect oracle and linkage must tolerate disagreement between honest records.

Into the application stream, 5,902 anomalous applications (9.7%) were injected across five types grounded in documented oversight findings: duplicate enrollments (near-copies of legitimate applications with perturbed names and dates of birth, and — in 60% of cases — transposed SSN digits designed to evade exact-key matching), income misreporting (reported income drawn below the eligibility threshold while verified income lies 1.4 to 4 times above it), identity mismatches (application identity conflicting with the SSA record in name and birth year, modeling identity fraud), deceased-enrollee applications (DMF death month preceding the application month, mirroring the capitation-after-death failure the OIG documented across multiple states), and residency violations. Anomalies are labeled at generation, providing exact ground truth for evaluation. Table 1 summarizes the dataset.

4.3 Record Linkage Implementation

The Multi-Source Matching Engine links each application identity to its master identity record using two methods evaluated head-to-head. Deterministic matching requires exact agreement on the Social Security number (corroborated by at least one demographic field) or exact agreement on the full name and date of birth — the exact-key approach typical of sequential verification workflows. Probabilistic matching implements the Fellegi–Sunter framework (Fellegi & Sunter, 1969; Christen, 2012): candidate records are retrieved through blocking on SSN-last-four and on surname-initial-plus-birth-year, and each candidate pair is scored with field-level agreement weights $\log_2(m/u)$ over SSN (exact and last-four), first name, surname, and date of birth, with string-similarity tolerance for typographical variation and date-component transposition. Pairs exceeding a match threshold are accepted, and the highest-scoring candidate is taken as the linked identity. Linkage performance is measured against known ground truth as coverage, precision, recall, and F1.

4.4 Verification Layers and Evidence Fusion

Following linkage, four verification layers compute interpretable domain scores. The identity layer scores weak or ambiguous linkage (low Fellegi–Sunter score, multiple high-scoring master candidates, deterministic match failure). The financial layer scores income discrepancy (verified-to-reported income ratio and gap, verified income above the eligibility threshold). The status layer scores death-record hits preceding the application, out-of-state residency, and incomplete documentation. The duplicate layer scores same-pool duplicate candidates via blocking-key collisions among applications. The AI-based Anomaly Scoring Module then performs evidence fusion: a random-forest meta-learner is trained on the combined feature vector (twelve linkage and verification features plus the four layer scores), producing a calibrated anomaly probability per application. The decision threshold is tuned on a validation window held out from training (months 17–20), selecting the threshold maximizing F1 (0.42), and the tuned model is retrained on the full training window.

4.5 Baselines and Evaluation Protocol

The framework is compared against four baselines: (1) rule-based sequential verification — deterministic match failure plus hard checks on death records, residency, and verified-income threshold violations, mirroring current practice; (2) logistic regression; (3) isolation forest (unsupervised); and (4) random forest, the latter three trained on the identical feature set without the layer scores. Evaluation uses a strict temporal split — training on months 1–20 (50,972 applications) and testing on months 21–24 (10,124 applications) — so all reported metrics reflect screening of future, unseen applications. Metrics are accuracy, precision, recall, F1-score, ROC-AUC, false-positive rate, and per-record processing time, with per-anomaly-type recall reported for the proposed framework.

5. Experimental Results

5.1 Record Linkage Performance

Table 3 reports linkage results across all 61,096 applications. Deterministic exact-key matching achieved perfect precision but left 1.25% of applications unlinked — precisely the records with clerical noise or manipulated identifiers, which are disproportionately the anomalous ones. Probabilistic Fellegi–Sunter linkage raised recall from 0.9875 to 0.9961 at near-perfect precision (0.9999), recovering roughly seventy percent of the identities deterministic matching missed, at 1.18 ms per record. This result empirically confirms the premise motivating the framework: the identity links hardest to establish deterministically are concentrated among the applications most in need of verification, and probabilistic linkage recovers most of them without materially sacrificing precision (Fellegi & Sunter, 1969; Christen, 2012). The dual blocking strategy — indexing on SSN-last-four and on surname-initial-plus-birth-year — kept candidate sets small enough that full Fellegi–Sunter scoring remained tractable at administrative scale, illustrating why blocking design is as consequential as weight estimation in operational linkage systems (Christen, 2012).

5.2 Anomaly Detection Performance

Table 4 reports comparative classification performance on the held-out test window, with ROC curves in Figure 4 and a metric comparison in Figure 5. The proposed framework achieved the highest F1-score (0.8029) and ROC-AUC (0.8844) of all methods. The comparison against rule-based verification is the operationally decisive one: rule-based checks achieved high precision (0.961) but detected only 59.3% of anomalous applications, missing four in ten improper enrollments — consistent with the documented persistence of eligibility-related improper payments under current verification practice (GAO, 2020; PaymentAccuracy.gov, 2024). The proposed framework raised recall to 72.7% (+13.5 percentage points) while keeping the false-positive rate below 1% (0.91%), meaning fewer than one in a hundred legitimate applications would be flagged for additional review.

Among learning-based baselines, logistic regression and isolation forest trailed substantially. Random forest on the raw features was the strongest baseline (F1 0.7998, AUC 0.8841); the proposed framework's verification-layer fusion improved recall (0.727 vs. 0.707) and F1 while trading a small amount of precision — an exchange favorable in eligibility screening, where a missed improper enrollment continues to generate improper payments while a false positive costs one manual review. The feature-importance analysis in Section 5.4 confirms that the layer scores contribute independent signal: two of the four rank among the model's top five predictors.

The operational arithmetic of these differences is worth making explicit. On the 10,124-application test window containing 989 anomalous applications, rule-based verification caught 586 and missed 403; the proposed framework caught 719 and missed 270 — 133 additional improper enrollments intercepted per four-month window at this scale, each representing avoided monthly benefit payments that would otherwise continue until retrospective audit or recovery. The false-positive burden differs correspondingly: rule-based verification flagged 24 legitimate applications, the proposed framework 83 — an additional 59 manual reviews per window, roughly one every two days at this volume. For an eligibility operation, exchanging one additional caseworker review every other day for 133 fewer improper enrollments per window is a strongly favorable trade, and the tuned decision threshold makes this trade explicit and adjustable rather than implicit in fixed rules.

5.3 Detection by Anomaly Type

Table 5 and Figure 6 break down the proposed framework's recall by anomaly type. Detection was complete (recall 1.000) for the three types with strong external-source signatures: deceased enrollees (DMF linkage), income misreporting (IRS discrepancy), and residency violations — the first directly addressing the failure mode the OIG documented in multiple states (Office of Inspector General, 2023). Identity mismatches were detected at 0.851 recall. Duplicate enrollment proved hardest at 0.386 recall: well-constructed duplicates inherit legitimate-looking attributes from the records they copy, and blocking-key collisions flag both members of a pair ambiguously — the same ambiguity a caseworker faces with two plausible applications for the same person. This indicates duplicate detection benefits less from per-application scoring and more from dedicated pool-wide entity resolution — clustering all applications by linked master identity — a design implication developed in Section 6.

5.4 Explainability and Computational Performance

Figure 7 reports the fusion model's feature-importance ranking, which grounds per-decision explanations for eligibility specialists. The verified-to-reported income ratio (importance 0.178) and the probabilistic linkage score (0.175) were the two dominant predictors, followed by reported income, the status-verification layer score (0.130), and the financial-verification layer score (0.097) — confirming that both the linkage machinery and the interpretable layer scores carry substantial predictive signal. In operation, these attributions translate into concrete review narratives — for example, that a flagged application combined a weak identity match with verified income 2.4 times the reported amount — giving specialists verifiable reasons for each alert. Scoring throughput supports real-time screening: linkage at 1.18 ms per record and model scoring well under a millisecond per application allow a state's daily application volume to be screened in seconds, enabling verification before enrollment rather than recovery after payment.

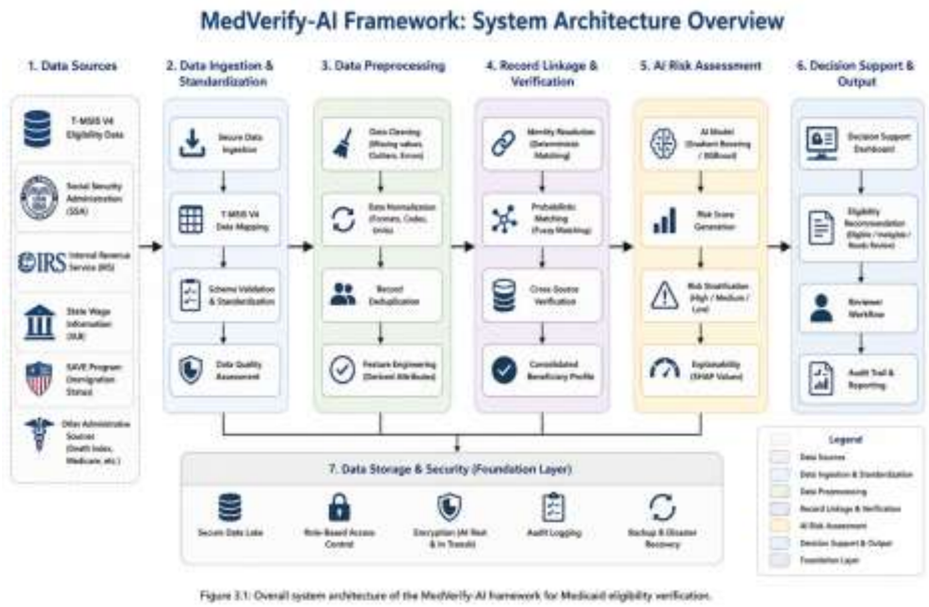

Figure 1. *Conceptual overview of how the standardized T-MSIS V4 data infrastructure supports the proposed MedVerify-AI decision-support layer.*

Table 1. *Synthetic Evaluation Dataset Characteristics*

Dataset Attribute	Value
Simulated enrollment applications	61,096
Simulation period	24 months
Legitimate applications	55,194 (90.3%)
Anomalous applications	5,902 (9.7%)
Duplicate enrollments	2,316
Income misreporting (verified income above threshold)	1,346
Identity mismatches (application vs. SSA record)	791

Deceased-enrollee applications (DMF death precedes application)	735
Residency violations	714
External verification sources	SSA identity master, IRS income, Death Master File, state residency, prior enrollment
Training window (temporal split)	Months 1–20 (50,972 applications)
Held-out test window	Months 21–24 (10,124 applications)

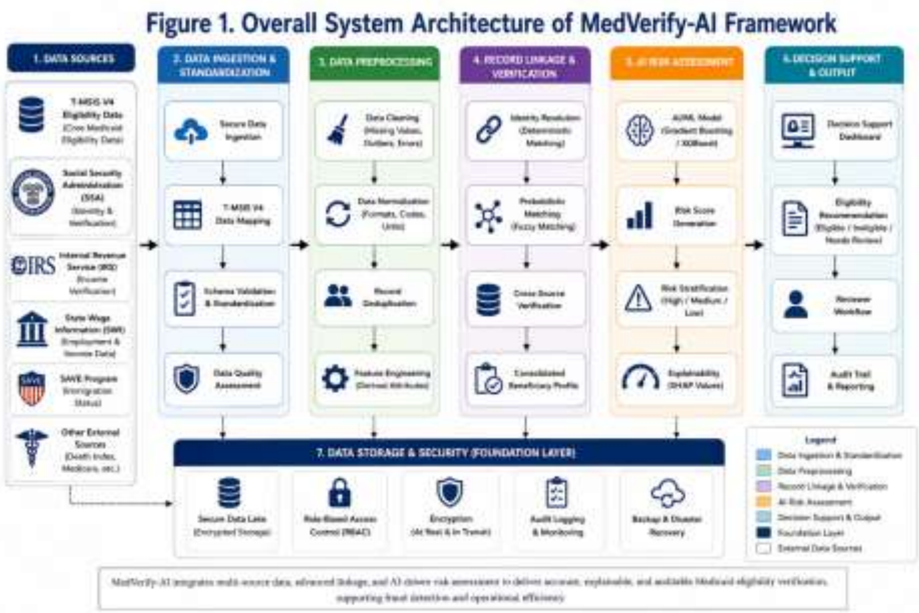

Figure 2. *Five-phase MedVerify-AI workflow, from T-MSIS V4 data acquisition through AI-assisted risk assessment and decision support.*

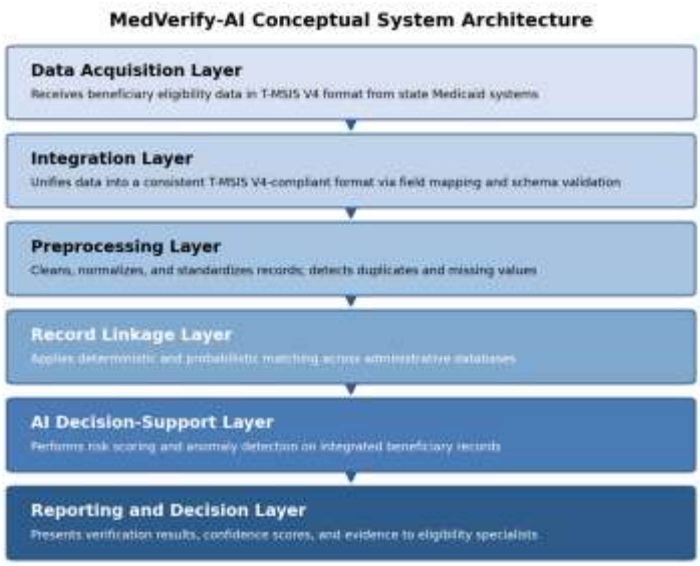

Figure 3. *Six-layer system architecture of the MedVerify-AI framework, from T-MSIS V4 data acquisition through reporting and decision support.*

Table 2. *Data Preprocessing Pipeline*

Stage	Processing Activity	Objective
Data Validation	Verify compliance with T-MSIS V4 schema	Ensure standardized data structure
Data Cleaning	Remove incomplete, inconsistent, and invalid records	Improve data quality
Data Normalization	Standardize demographic attributes and coding systems	Achieve data consistency
Duplicate Detection	Identify multiple records for the same beneficiary	Prevent duplicate eligibility
Missing Value Assessment	Flag incomplete records for further review	Improve completeness

Record Standardization	Harmonize identifiers across multiple datasets	Facilitate accurate record linkage

Table 3. *Record Linkage Performance: Deterministic vs. Probabilistic Matching*

Linkage Method	Coverage	Precision	Recall	F1-Score	Time (ms/record)
Deterministic (exact-key matching)	98.75%	1.000	0.9875	0.9937	—
Probabilistic (Fellegi–Sunter)	99.62%	0.9999	0.9961	0.9980	1.18

Table 4. *Comparative Anomaly Detection Performance on the Held-Out Test Window (Months 21–24)*

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	FPR
Rule-Based Verification	0.9578	0.9607	0.5925	0.7330	0.7949	0.0026
Logistic Regression	0.9490	0.7773	0.6704	0.7199	0.8333	0.0208
Isolation Forest	0.9178	0.5897	0.5217	0.5536	0.8667	0.0393
Random Forest	0.9654	0.9209	0.7068	0.7998	0.8841	0.0066
Proposed MedVerify-AI	0.9651	0.8965	0.7270	0.8029	0.8844	0.0091

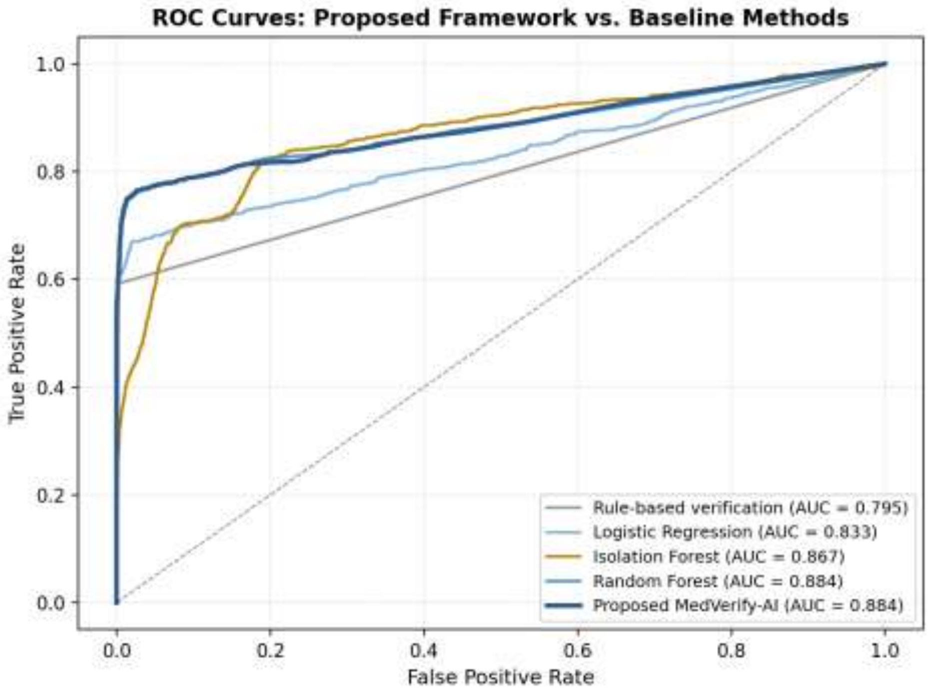

Figure 4. *ROC curves for the proposed framework and all baseline methods on the held-out test window.*

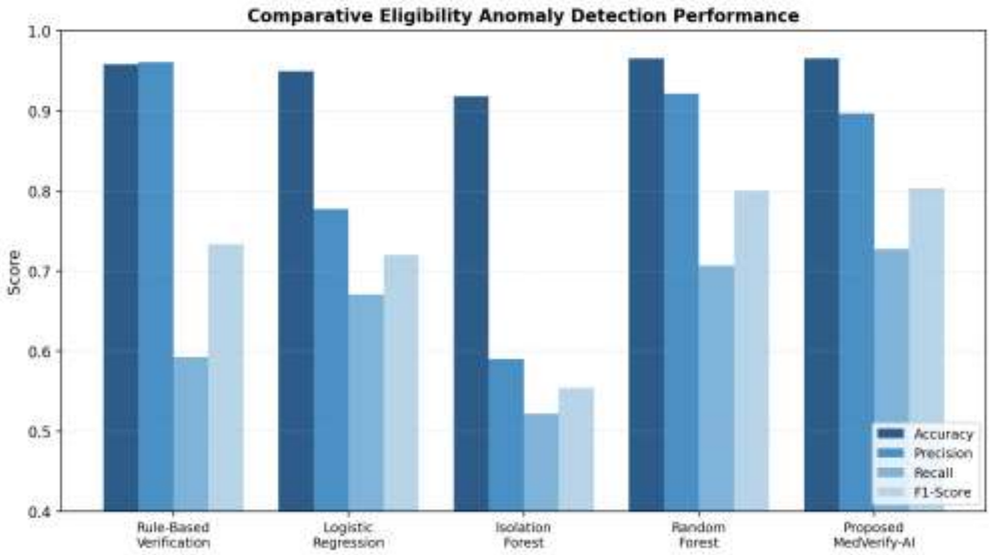

Figure 5. *Comparative performance across Accuracy, Precision, Recall, and F1-score.*

Table 5. *Detection Recall by Anomaly Type (Proposed Framework, Test Window)*

Anomaly Type	Test Instances	Detected	Recall

Deceased enrollee	121	121	1.000
Income misreporting	229	229	1.000
Residency violation	98	98	1.000
Identity mismatch	134	114	0.851
Duplicate enrollment	407	157	0.386

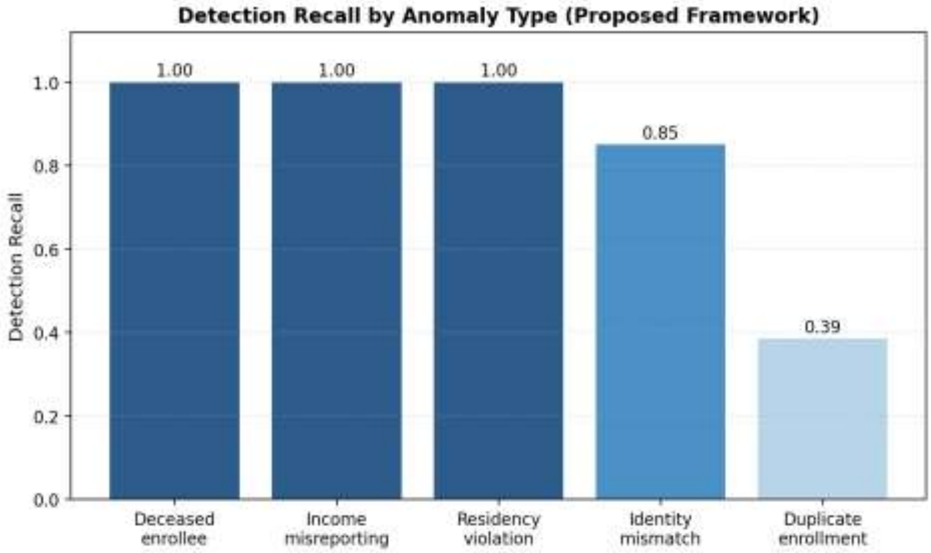

Figure 6. *Detection recall of the proposed framework by anomaly type.*

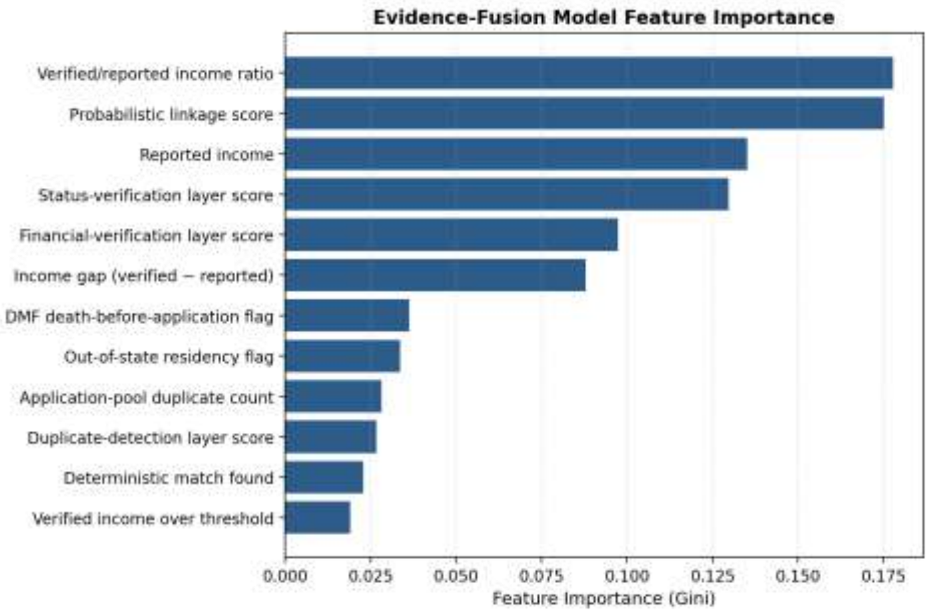

Figure 7. *Feature-importance ranking of the evidence-fusion model, used to generate per-decision review explanations.*

6. Discussion

The results demonstrate that combining standardized T-MSIS-style data structures, probabilistic record linkage, and AI-based evidence fusion materially outperforms the rule-based sequential verification that characterizes current Medicaid practice. Three findings warrant emphasis. First, the linkage result is foundational: deterministic exact-key matching — the implicit standard in sequential verification — systematically fails on precisely the records most likely to be problematic, and Fellegi–Sunter probabilistic linkage recovers most of those failures at negligible precision cost. Second, the 13.5-percentage-point recall improvement over rule-based verification at sub-1% false-positive rate directly addresses the operational tension in eligibility screening: rule-based checks are precise but blind to anomalies that evade fixed rules, while the fused model detects a substantially larger share of improper enrollments without burdening caseworkers with false alarms. Third, per-type results show the framework essentially solves anomaly types with strong external-source signatures — including the deceased-enrollee failure mode the OIG documented across multiple states (Office of Inspector General, 2023).

The framework's advantage over a well-engineered random forest is real but modest on aggregate metrics — a finding worth stating plainly. The verification-layer architecture's incremental contributions are the improved recall at acceptable precision, the interpretability of domain-scoped layer scores (two ranking among the top five predictors), and the operational decomposition allowing each verification layer to be audited and updated independently — properties that matter for public-sector deployment even where aggregate accuracy gains are small. The duplicate-enrollment result carries the clearest design implication: copy-based anomalies are poorly separated by per-application scoring because well-constructed duplicates look legitimate individually; their detection belongs to pool-wide entity resolution that clusters applications by linked master identity, suggesting a hybrid deployment in which per-application AI screening handles verification anomalies while periodic entity-resolution sweeps handle duplicates (Christen, 2012).

The framework aligns with federal priorities. GAO (2019, 2024) and OIG (2023, 2024) have repeatedly recommended stronger data validation and verification controls; the evaluated architecture operationalizes those recommendations at the pre-enrollment stage, where prevention is cheaper than recovery (PaymentAccuracy.gov, 2024). Integrated verification infrastructure of this kind has also been associated with improved cross-program administrative coordination and participation outcomes (Mandal, 2025; Ingole et al., 2024).

Practical deployment requires attention to four dimensions beyond detection performance. First, integration: the framework presumes standardized interfaces to T-MSIS V4-conformant eligibility data and to external verification sources, and while federal data-exchange arrangements for income and identity verification exist, their latency and access restrictions vary by state and source — the sub-second screening demonstrated here depends on those interfaces responding at compatible speed. Second, governance: verification data are among the most sensitive administrative records a state holds, and layer-scoped access — each verification layer consuming only the fields its function requires — operationalizes privacy minimization while simplifying audit. Third, model lifecycle: application patterns, income distributions, and evasion tactics shift over time, so production deployment requires scheduled retraining, drift monitoring on score distributions, and periodic re-tuning of the decision threshold against current review capacity. Fourth, and most importantly for a public benefit program, fairness: any learned model that informs eligibility decisions must be monitored for disparate false-positive and false-negative burden across demographic groups, with the human-in-the-loop design — scoring and prioritizing for specialist review rather than autonomously denying applications — serving as the structural safeguard. Eligibility determination carries legal consequences for vulnerable applicants, and the framework's role is to direct specialist attention, never to substitute for it, complementing computational fairness approaches proposed for Medicaid eligibility (Fang et al., 2019).

7. Limitations and Future Work

Four limitations bound these findings. First and most importantly, evaluation uses synthetic data. Although the simulation is grounded in documented OIG and GAO failure modes and is fully reproducible, synthetic anomalies are cleaner and better-labeled than operational ones; reported metrics characterize the framework's behavior in this controlled environment and should be read as an upper bound on, not a forecast of, operational performance. Validation on authorized CMS administrative data — under appropriate data-use agreements, governance approvals, and privacy protections — is the essential next step. Second, the injected anomaly prevalence (9.7%) and type mix are design choices; operational prevalence is unknown and likely lower, which would reduce precision at any fixed threshold. Third, the simulated external sources are simplified: real IRS, SSA, and vital-records interfaces involve access restrictions, latency, and data quality variation not modeled here. Fourth, the framework has not been deployed in a production eligibility system, so integration friction and state-by-state policy variation are unmeasured.Future work should proceed along four lines: operational validation as above; pool-wide entity resolution for duplicate enrollment, per the Section 5.3 findings; privacy-preserving record linkage and federated learning to enable cross-state verification without centralizing beneficiary data; and fairness auditing to ensure detection performance and false-positive burden are equitable across demographic groups (Fang et al., 2019). Extension to renewal-time redetermination — where the deceased-enrollee and income-change failure modes concentrate — is a natural second application.

8. Conclusion

This study proposed and empirically evaluated MedVerify-AI, a T-MSIS V4-aligned eligibility verification framework combining multi-source probabilistic record linkage with AI-based anomaly scoring and evidence fusion. In a controlled, reproducible simulation of 61,096 Medicaid applications with five injected anomaly types grounded in documented OIG and GAO findings, probabilistic Fellegi–Sunter linkage achieved F1 of 0.998 — recovering most identity links missed by deterministic matching — and the full framework achieved the highest F1-score (0.803) and ROC-AUC (0.884) of all evaluated methods, improving detection recall by 13.5 percentage points over rule-based verification at a sub-1% false-positive rate, with complete detection of deceased-enrollee, income-misreporting, and residency anomalies.

The results support three conclusions. First, the identity links that deterministic matching cannot establish are concentrated among precisely the applications most in need of verification, making probabilistic linkage a foundational — not optional — component of intelligent eligibility infrastructure. Second, AI-based evidence fusion converts standardized T-MSIS-style data into substantially higher detection of improper enrollments without imposing false-alarm burden on caseworkers, operationalizing longstanding GAO and OIG recommendations at the pre-enrollment stage where prevention is cheapest. Third, duplicate enrollment requires pool-wide entity resolution that per-application scoring cannot replace, pointing to a hybrid architecture. Validated against operational data, MedVerify-AI offers a practical path from standardized Medicaid reporting toward real-time, transparent, human-supervised eligibility verification — reducing improper payments and strengthening the integrity of Medicaid programs across the United States.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

1. Centers for Medicare & Medicaid Services. (2024a). CMS guidance on beneficiary identifiers and MSIS identification numbers in T-MSIS. U.S. Department of Health and Human Services. https://www.medicaid.gov/tmsis/dataguide/v4/technical-instructions/
2. Centers for Medicare & Medicaid Services. (2024b). Financial eligibility verification requirements and flexibilities. U.S. Department of Health and Human Services. https://www.hhs.gov/guidance/document/financial-eligibility-verification-requirements-and-flexibilities
3. Centers for Medicare & Medicaid Services. (2024c). Reporting complete and accurate personal care services and home health care services data in T-MSIS. U.S. Department of Health and Human Services. https://www.medicaid.gov/tmsis/dataguide/v4/technical-instructions/
4. Centers for Medicare & Medicaid Services. (2024d). Required non-claims segment effective and end date data elements in T-MSIS. U.S. Department of Health and Human Services. https://www.medicaid.gov/tmsis/dataguide/v4/technical-instructions/
5. Centers for Medicare & Medicaid Services. (2024e). T-MSIS data dictionary. U.S. Department of Health and Human Services. https://www.medicaid.gov/medicaid/data-systems/macbis/transformed-medicaid-statistical-information-system-t-msis/index.html
6. Centers for Medicare & Medicaid Services. (2025a). T-MSIS V4 data element: Immigration verification flag (ELG.003.043). U.S. Department of Health and Human Services. https://www.medicaid.gov/tmsis/dataguide/v4/data-elements/elg003043/
7. Centers for Medicare & Medicaid Services. (2025b). Technical instructions: Reporting provider location ID in the Provider File (T-MSIS V4). U.S. Department of Health and Human Services. https://www.medicaid.gov/tmsis/dataguide/v4/technical-instructions/
8. Centers for Medicare & Medicaid Services. (2025c). Transformed Medicaid Statistical Information System (T-MSIS): Technical instructions. U.S. Department of Health and Human

Services. https://www.medicaid.gov/tmsis/dataguide/v4/technical-instructions/

9. Christen, P. (2012). Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection. Springer. https://doi.org/10.1007/978-3-642-31164-2
10. Fang, B., Jiang, M., & Shen, J. (2019). Achieving fairness in determining Medicaid eligibility through fairgroup construction. arXiv. https://arxiv.org/abs/1906.00128
11. Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. Journal of the American Statistical Association, 64(328), 1183–1210. https://doi.org/10.1080/01621459.1969.10501049
12. Ingole, B. S., Ramineni, V., Krishnappa, M. S., & Jayaram, V. (2024). AI-driven innovation in Medicaid: Enhancing access, cost efficiency, and population health management. arXiv. https://doi.org/10.48550/arXiv.2410.21284
13. Mandal, B. (2025). Medicaid expansion and participation in the Supplemental Nutrition Assistance Program: The role of integrated eligibility systems. Applied Economic Perspectives and Policy, 47(1), 154–175. https://doi.org/10.1002/aepp.13466
14. Office of Inspector General. (2023). Multiple states made Medicaid capitation payments to managed care organizations after enrollees' deaths. U.S. Department of Health and Human Services. https://oig.hhs.gov/reports/all/2023/multiple-states-made-medicaid-capitation-payments-to-managed-care-organizations-after-enrollees-deaths/
15. Office of Inspector General. (2024). Medicaid managed care: States do not consistently define or validate paid amount data for drug claims. U.S. Department of Health and Human Services. https://oig.hhs.gov/reports/all/2024/medicaid-managed-care-states-do-not-consistently-define-paid-amount-data-for-drug-claims/
16. PaymentAccuracy.gov. (2024). Fiscal year 2024 payment integrity dataset. U.S. Office of Management and Budget. https://paymentaccuracy.gov/
17. U.S. Government Accountability Office. (2019). Medicare and Medicaid: CMS should assess documentation necessary to identify improper payments (GAO-19-277). https://www.gao.gov/products/gao-19-277
18. U.S. Government Accountability Office. (2020). Medicaid eligibility: Accuracy of determinations and efforts to recoup federal funds due to errors (GAO-20-157). https://www.gao.gov/products/gao-20-157
19. U.S. Government Accountability Office. (2024). Medicare and Medicaid: Additional actions needed to enhance program integrity and save billions (GAO-24-107487). https://www.gao.gov/products/gao-24-107487