Integrating Machine Learning for Health Finance Analytics: Predicting Healthcare Costs and Economic Outcomes in the United States

Sophia L Carter [1*], **Michael T Reynolds** [2], **Lauren A Bennett** [3]
[1] University of Michigan, Ann Arbor, MI, United States
[2-3] Southern Illinois University Edwardsville, Edwardsville, IL, United States

* Corresponding Author: **Sophia L Carter**

Article Info

P-ISSN: 3051-3502
E-ISSN: 3051-3510
Volume: 06
Issue: 02
Received: 27-10-2025
Accepted: 29-11-2025
Published: 24-12-2025
Page No: 209-217

Abstract
Machine learning can do more than forecast claims. It can connect clinical risk, social context, and macroeconomic signals to predict what patients, payers, and regional economies will actually face. This paper proposes and evaluates a U.S. health finance analytics blueprint that joins person-level utilization, community-level social determinants, and national indicators to predict annual healthcare costs and economic outcomes. We integrate widely used surveys and administrative datasets, harmonize coding, and align time windows so that predictions aggregate cleanly from people to places. Methodologically, we use a two-part framework for spending (any spend classifier plus conditional cost regressor), gradient boosted trees for nonlinear interactions, and explainable AI with Shapley values to expose drivers of cost and financial risk. We pair individual predictions with county level forecasts of outcomes relevant to households and governments, including out of pocket exposure, medical debt risk, and employment impacts. We also embed fairness tests, privacy safeguards, and model governance consistent with health data regulations.To illustrate feasibility, we present a reproducible pipeline with model comparison, error analysis, and interpretable feature summaries. The approach targets practical decisions: pricing and benefit design for insurers, budgeting and value-based payment for providers, and policy planning for agencies. Results from a worked example (synthetic for demonstration) show that gradient boosting improves explanatory power over linear baselines while remaining interpretable, and that adding social and economic context reduces error for high need patients. We conclude with implementation guidance, limitations, and a research agenda to validate this integrated health finance framework on public microdata and to study its equity and policy implications at scale.

DOI: https://doi.org/10.54660/IJMER.2025.6.2.209-217

Keywords: Health Finance Analytics, Healthcare Expenditures, Cost Prediction, Machine Learning, Gradient-Boosted Trees (XGBoost), Two-Part Models, Explainable AI (SHAP), Algorithmic Fairness, Social Determinants of health (SDOH), Out-of-Pocket Costs, Medical Debt, United States

1. Introduction

Healthcare costs shape U.S. household budgets, employer decisions, and public finances. When spending rises faster than the economy, trade-offs emerge: premiums crowd out wages, state budgets tighten, and families delay care. Recent national accounts show health spending of about $4.9 trillion in 2023 up 7.5 percent while coverage and payer mix shifted as the public health emergency ended. Those facts underscore a simple point: without reliable, explainable prediction, stakeholders fly blind. Accurate forecasts of who will incur high costs and where budgets will strain are now essential to pricing, benefit design, and policy. (Centers for Medicare & Medicaid Services, 2024) [4].

Here's the thing: we already collect the raw ingredients for better forecasting. Person level utilization and payments appear in the Medical Expenditure Panel Survey (MEPS), and hospital encounters are captured in AHRQ's Healthcare Cost and Utilization Project (HCUP). Local social context from poverty and unemployment to housing cost burden is available via CDC's PLACES project and the American Community Survey. Each source spans a different level of analysis, but together they trace how clinical risk, social determinants, and insurance combine to produce spending. The opportunity is to harmonize these layers and deliver predictions that aggregate cleanly from people to places to national accounts. (AHRQ MEPS, 2019; AHRQ HCUP, n.d.; CDC PLACES, 2024; U.S. Census Bureau, 2025) [1, 2, 27].

Policy has nudged the data pipes wider. The 21st Century Cures Act interoperability rule, TEFCA's common agreement, and widespread adoption of HL7 FHIR APIs pushed information blocking prohibitions and baseline technical standards into daily practice. That broader access raises the bar for responsible analytics. Any model that influences coverage, pricing, or care management needs transparent reasoning and rigorous privacy practices under HIPAA's de identification guidance. Put simply, better plumbing demands better governance. (ONC, 2020; HHS Office for Civil Rights, 2025; ONC/Sequoia RCE, 2023; HL7 FHIR Foundation, 2025) [18, 19, 14].

Machine learning can meet this moment if we prove three things. First, that integrated models beat strong baselines at predicting individual and regional costs without turning into black boxes. Second, that predictions scale up to economic outcomes people actually feel: out of pocket exposure, medical debt risk, and employment impacts. Third, that fairness is addressed up front. Predicting "future spend" can encode structural inequities when dollars are used as a proxy for need; a well-known study showed exactly that dynamic in a commercial population health algorithm. Our framework therefore separates clinical need from observed spend, audits predictions with equalized odds metrics, and reports feature contributions with model agnostic explanations. (Obermeyer, Powers, Vogeli, & Mullainathan, 2019; Hardt, Price, & Srebro, 2016; Lundberg & Lee, 2017) [20, 8, 16].

Why integrate finance explicitly? Because the costs we predict feed directly into household and macro indicators. KFF and Urban Institute work highlights the persistence of out of pocket burdens and the geography of medical debt, even as credit reporting rules evolve. A joined-up model lets leaders test how benefits, provider prices, or economic shocks translate into patient exposure and local debt risk. This is where health finance analytics lives: at the interface between clinical utilization and real-world affordability. (KFF, 2024; Urban Institute, 2023) [15, 28].

Finally, we situate the work in macro context. Real GDP grew, yet health spending expanded faster, widening the share of the economy devoted to care. That gap drives employer premium increases and pressures state Medicaid budgets, even as balance billing protections under the No Surprises Act shift some burden from patients. Any credible cost prediction system should show how micro predictions roll up to payer budgets and how policy levers alter out of pocket exposure. Our approach makes those links testable. (BEA, 2024; CMS, 2024; CMS No Surprises Act, 2025) [24, 4, 5]. Clearly.

Literature Review

Predicting healthcare spending has a long history in health services research and actuarial science. Classical approaches include generalized linear models with log links and gamma errors, as well as two part models that separate the probability of any expenditure from the conditional amount among spenders. In the last decade, studies have tested random forests, gradient boosting, and deep learning on large claims to identify high cost patients and forecast total spend. Systematic and scoping reviews find that machine learning can improve discrimination and reduce error over linear baselines, although performance depends on data breadth and careful validation; weak feature engineering and leakage can erase apparent gains. Empirical work shows competitive results for boosting and neural networks in predicting next year spend or high cost utilization, while cautioning that interpretability and shift robustness remain open challenges. (Huang et al., 2022; BMC MI&D, 2021; PLOS ONE, 2023; MDPI, 2024; Data, 2025) [7, 32, 31, 30, 29].

Datasets suitable for U.S. cost prediction are mature but fragmented. MEPS offers nationally representative person level microdata on utilization, payments, insurance, health status, and demographics, making it ideal for individual prediction and fairness audits. HCUP's state and national files add encounter level detail on diagnoses, procedures, charges, and outcomes; they support hospital and market level modeling of costs and utilization. For community context, CDC's PLACES provide small area estimates of chronic disease, risk behaviors, and social determinants, while the Census Bureau's ACS and CPS ASEC provide health insurance coverage, income, and poverty. Linking these sources creates a multilevel view of how clinical risk and context combine to produce spending. (AHRQ MEPS, 2019; HCUP, n.d.; CDC PLACES, 2024; Census, 2025) [1, 2, 27].

A third pillar is the national accounting of health spending and macroeconomy indicators that define the fiscal backdrop for prediction. CMS's National Health Expenditure (NHE) series documents the level and growth of spending by payer and service; the most recent release reports $4.9 trillion in 2023 with growth of 7.5 percent. Independent coverage reached similar conclusions and underscored the compositional shifts as the public health emergency unwound. BEA's GDP releases, BLS's CPI and unemployment series, and Census coverage statistics round out the macro context and provide covariates for economic outcome models. (CMS, 2024; Health Affairs, 2024; BEA, 2024; BLS, 2024; Census, 2025) [4, 24, 25, 26, 27].

Interoperability and regulation shape what is feasible. The 21st Century Cures Act final rule prohibits information blocking and drives standardized APIs into certified EHRs. TEFCA's common agreement establishes baseline policy and technical terms for cross network exchange; ONC issued updates through 2024–2025 to operationalize the framework. HL7's FHIR specification provides resource models and RESTful patterns that make it easier to extract, normalize, and link clinical data to claims and community indicators. HIPAA's de identification guidance outlines Safe Harbor and Expert Determination pathways that are compatible with linkage projects when handled correctly. Together these rules and standards have shifted the bottleneck from access to governance and quality. (ONC, 2020; ONC TEFCA Final

Rule, 2025; HL7 FHIR, 2025; HHS OCR, 2025) [18, 14, 19].

Methodologically, three threads stand out. First, gradient boosted trees such as XGBoost have become a workhorse for tabular health finance problems because they capture nonlinear interactions, handle missingness well, and remain relatively interpretable with post hoc tools. Second, L1 regularized linear models (lasso) provide sparse baselines that are easy to explain and can perform strongly when effects are close to linear. Third, additive time series models such as Prophet support transparent calendar effects and can be combined with machine learned residuals for budget forecasting. (Chen & Guestrin, 2016; Tibshirani, 1996; Taylor & Letham, 2018) [6, 23, 22].

Explainability and fairness are now center stage. SHAP connects complex models to additive feature attributions with desirable theoretical properties, making it possible to decompose patient level cost predictions into drivers such as prior year spend, multimorbidity, medication burden, and neighborhood deprivation. Equality of opportunity and related criteria offer testable definitions of nondiscrimination in supervised learning. Evidence from health systems shows why these tests matter: a widely used commercial algorithm optimized predicted cost and thereby under identified Black patients for extra services relative to equally sick White patients. Fixing the target to predicted need reduced the bias. (Lundberg & Lee, 2017; Hardt, Price, & Srebro, 2016; Obermeyer, Powers, Vogeli, & Mullainathan, 2019) [16, 8, 20].

Work specifically pitched at “health finance analytics” is growing. KFF’s Health System Tracker synthesizes spending, prices, utilization, and out of pocket burdens across payers, and shows that cost sharing remains a consequential slice of spending for people with employer coverage. Urban Institute’s medical debt analyses map debt in collections at the county level and show how coverage and income dynamics affect exposure. On the policy side, the No Surprises Act created federal protections against out of network balance bills; implementation documents are now detailed enough to model which services and plan types fall within those protections. Together, these sources support county and plan level dependent variables that align with lived economic outcomes. (KFF, 2024; Urban Institute, 2023; CMS No Surprises Act, 2025) [15, 28, 5].

Adjacent literatures reinforce the integrated framing. In oncology, OncoViz demonstrates how model driven dashboards can surface incidence, screening, and mortality disparities with an eye toward targeted investment and policy (Hasan, Bhuyain, Chowdhury, & Arman, 2021) [10]. In health operations, supply chain optimization research details how predictive analytics improves resilience and reduces waste lessons that translate directly to capacity planning and inventory for high cost therapies (Rasel, Arman, Hasan, & Bhuyain, 2022) [21]. Financial security work argues for AI driven fraud detection and risk analytics across payments and claims, a design pattern that healthcare can adopt when building real time monitoring for anomalous billing and identity misuse (Hasan *et al*., 2025b) [12]. PRISMA reviews of predictive analytics for financial information security, and broader cybercrime syntheses, further codify methods supervised and graph based anomaly detection, stream modeling, and adversarial testing that port cleanly to payer operations. (Hasan *et al*., 2025a; Milon *et al*., 2024) [11, 17].

Several U.S. focused contributions explicitly tie predictive analytics to cost control and improved outcomes. Hasan and colleagues survey strategies for cost reduction and outcome improvement using predictive analytics in U.S. healthcare and emphasize deploying interpretable models and domain specific dashboards to drive operational decisions (Hasan, Arman, Bhuyain, Chowdhury, & Bathula, 2025) [9]. Related energy systems work by the same group highlights how big data forecasting under uncertainty supports transition planning at national scale; the methodological logic (forecast, explain, act) carries over to health budgets (Arman, Hasan, & Rasel, 2024) [3]. Security oriented papers stress the national security stakes of protecting health data and the importance of machine learning defenses in healthcare infrastructure (Hasan *et al*., 2022) [13]. Taken together, these studies make the case for pipelinesthat are predictive, interpretable, and governable.

In sum, the literature suggests four design principles. First, predict costs with methods that capture nonlinearity but remain amenable to explanation boosted trees with post hoc attributions or sparse linear baselines. Second, add context: link person level utilization to neighborhood deprivation, local labor markets, and insurance coverage. Third, evaluate models not only on error but also on fairness and policy relevance does the model change out-of-pocket exposure or shift budget risk in a way consistent with statutory protections such as the No Surprises Act? Fourth, treat privacy, security, and interoperability as first-class concerns; the policy stack now enables data flow, which raises the accountability bar for analytics in U.S. settings.

Methodology

Design overview. We build and evaluate an integrated health finance pipeline that predicts (a) annual person level spend, (b) plan and county level aggregates of out of pocket exposure and medical debt risk, and (c) macro linked indicators such as employment effects. The design follows three principles: link the right levels of data, separate modeling tasks that have different error structures, and keep explanations and fairness checks first class. The pipeline is reproducible and intended for deployment in a payer or public health analytics environment.

Data sources and unit of analysis. Person level features come from MEPS household files (utilization, expenditures, insurance, health status, demographics) joined to longitudinal panels. Encounter level features come from HCUP state or national files (diagnoses, procedures, charges, discharge disposition). Community features come from CDC PLACES (health outcomes, behaviors, and social determinants), the American Community Survey (insurance coverage, income, poverty), and labor market indicators from the Bureau of Labor Statistics. Macro context for model to account reconciliation uses CMS NHE and BEA GDP releases. Person level observations are aligned to calendar years, with look back windows restricted to information available by December 31 of the prediction year 1. (MEPS; HCUP; CDC PLACES; ACS/CPS; BLS; CMS NHE; BEA) [1, 2, 26, 4, 24].

Cohorts and outcomes. We define two primary outcomes: (1) total annual expenditures per person (all payers), and (2) an out of pocket exposure proxy defined as cost sharing plus non covered spending. Secondary county level outcomes include the share of adults with medical debt in collections (from credit panel research) and the unemployment rate. For illustration in the figures we provide a synthetic worked example; the code and variable definitions are the same as those used for public microdata. (Urban Institute; BLS) [28, 26].

Ethics, privacy, and interoperability. Person level files are de identified according to HIPAA guidance using Safe Harbor or Expert Determination, depending on the source and linkage strategy. We retain only the minimum necessary fields for modeling and reporting. When integrating with provider data, we rely on ONC's Cures Act rules and TEFCA participation to streamline access with appropriate governance and auditability, and we export clinically relevant features via HL7 FHIR resources where possible. (HHS OCR, 2025; ONC, 2020; TEFCA Final Rule, 2025; HL7 FHIR) [19, 18, 14].

Feature engineering. We compute features across five groups:

1. **Utilization and severity:** prior year spend; counts of inpatient, ED, and outpatient visits; procedure classes; diagnosis groups; comorbidity indices; medication counts by therapeutic class.
2. **Insurance and access:** coverage type, continuity, high deductible plan indicator, usual source of care, delays due to cost.
3. **Community context:** PLACES SDOH measures (housing cost burden, crowding, single parent households), county poverty and unemployment, coverage rates, rurality.
4. **Prices and payment:** local price indices when available; payer mix.
5. **Policy flags:** service categories within No Surprises Act protections.

All continuous features are winsorized at the 99th percentile; counts are transformed with log(1+x). Categorical variables are encoded with target frequency smoothing to avoid high cardinality overfitting. (CDC PLACES; CMS No Surprises Act) [5].

Modeling strategy for spend. Because expenditures are zero inflated and right skewed, we use a two-part approach: a classifier for any spend vs no spend, then a regressor for positive spend. For the classifier we compare logistic regression, random forest, and gradient boosting. For the regressor we compare lasso regularized linear models, Tweedie GLMs, and gradient boosted trees (XGBoost). We fit both parts on the same training folds and multiply probabilities by conditional predictions to produce an expected spend. Hyperparameters are tuned with nested rolling origin cross validation; early stopping is used for boosting. (Tibshirani, 1996; Chen & Guestrin, 2016) [23, 6].

Aggregation and economic linkage. Individual predictions are aggregated to plan and county levels using survey weights where appropriate. We then estimate county level models that map predicted spend, coverage, and SDOH features to outcomes such as medical debt prevalence and unemployment. For budget projections, we combine plan level expected spend with additive time series forecasts that capture seasonality and policy events; residual structure is modeled with gradient boosting. Scenario covariates (e.g., a 1 percentage point unemployment shock) are passed through both the time series and cross sectional models to produce consistent projections. (Taylor & Letham, 2018) [22].

Interpretability

All tree-based models are accompanied by post hoc explanations using SHAP values to quantify global feature importance and local contributions for individual predictions. For linear models we report standardized coefficients. We aggregate SHAP values by clinically meaningful groups (multimorbidity, utilization intensity, SDOH) to align explanations with operational levers. We also provide "what if" profiles that hold clinical features fixed while varying community variables to quantify the sensitivity of predictions to social context. (Lundberg & Lee, 2017) [16].

Fairness and target choice. We evaluate equalized odds differences for the any spend classifier and compare calibration curves and absolute error by race/ethnicity and income. Because optimizing on dollars can encode inequities, we test an alternative target that approximates clinical need (e.g., predicted service volume or risk adjusted utilization) and compare disparities with the spend based target. When materially different, we prefer the need based target for allocation decisions and report spend only as a budget consequence. Threshold adjustments are explored but not used if they lower sensitivity for high need groups. (Hardt, Price, & Srebro, 2016; Obermeyer et al., 2019) [8, 20].

Evaluation metrics

For spend, we report MAE, RMSE, and outlier robust mean absolute percentage error; for classification, AUROC, AUPRC, and calibration (Brier score, reliability curves). For aggregates we report R^2 on county level outcomes and coverage of prediction intervals from the time series forecasts. All metrics are reported overall and by subgroups. We include ablation studies that remove SDOH and labor market features to quantify their incremental contribution. Synthetic figures show the expected pattern: boosting outperforms sparse linear baselines on R^2 while remaining explainable. (Figures 2–3; Supplementary Table).

Governance, monitoring, and security. We log every model artifact, data snapshot, and hyperparameter configuration; we version the data transformations and explanation dashboards. Drift monitoring covers covariate shift, calibration drift, subgroup error differentials, and changes in the distribution of explanations. We integrate fraud signal checks into the same pipeline graph-based provider anomalies and duplicate billing so that cost prediction and payment integrity share infrastructure. Security controls follow industry guidance on de identification and minimum necessary data, and access is brokered via FHIR APIs and TEFCA where possible. (HHS OCR, 2025; HL7 FHIR; ONC TEFCA) [19, 14, 18].

Reproducibility and software. Models are implemented in open source Python toolchains with feature registries, unit tests for feature logic, and CI linked to data snapshots. Artifacts models, hyperparameters, calibration curves and fairness reports are versioned. We provide scripts to rebuild public data cohorts and to export clinician explanations. The figures in this paper were generated by scripts with synthetic inputs so readers can reproduce this workflow without accessing protected health information.

Missing data and weighting. MEPS microdata include survey weights and complex design features. We use person level weights for national models and re weight to plan demographics when needed. Missingness is handled with indicator augmented median imputation for continuous fields and explicit "unknown" levels for categoricals; missing value flags are retained as features. HCUP encounters are linked to persons using consistent pseudo identifiers or probabilistic linkage on facility, date, and demographics. (HCUP documentation).

Entity resolution and coding harmonization. Diagnoses and procedures are mapped to stable clinical groups (e.g., CCS or HCC) to reduce sparsity. We normalize price fields by local market baskets where available and reconcile service dates across sources. Transformations are scripted and versioned for auditability. (CMS Risk Adjustment; HCUP).

Risk adjustment features. We compute CMS HCC prospective risk scores using the latest mappings and track the 2024–2025 phase in rules so plan revenue expectations align with model outputs. Scores are features, not targets; we also predict risk adjusted utilization to separate morbidity from prices. (CMS training materials, CY2025).

Training and tuning. We use rolling windows: train on t 4…t 1, validate at t, test at t+1. Hyperparameters are tuned with Bayesian optimization; class imbalance is handled with calibrated probabilities and cost sensitive thresholds. County level models include light spatial smoothing to stabilize low population areas.

Calibration and post processing. Expected spend is calibrated with isotonic regression per subgroup and constrained so that plan aggregates match top down totals within confidence bounds. Before deployment for care management or benefit design, we run policy stress tests that simulate subgroup inclusion rates and out of pocket exposure under No Surprises Act protections. (CMS No Surprises Act) [5].

Secure deployment. Identified data remain within the covered entity; de identified feature vectors feed the modeling service behind network segmentation. FHIR Subscriptions drive incremental updates without moving raw notes. Access logs, differential privacy budgets for ad hoc queries, and periodic red team exercises are standard. (HL7 FHIR; HHS OCR, 2025) [14, 19].

Outputs and decision products. The system emits (1) individual predictions with explanations; (2) plan and county level dashboards of projected spend, out of pocket exposure, and medical debt risk; and (3) an audit packet listing data sources, feature logic, fairness diagnostics, and calibration results. Figures 2–3 visualize performance and drivers in a synthetic but faithful workflow.

Discussion

Our worked example synthetic but aligned with public data structure points to three practical conclusions. First, integrated models that fuse clinical history with social and economic context reduce error relative to sparse linear baselines. In our illustration, gradient boosting delivered noticeably higher R² than lasso or plain linear regression (Figure 2), while remaining interpretable through additive explanations (Figure 3). Second, the same feature families that drive spend also connect to household financial exposure. Prior year spend, multimorbidity, and benefit design variables dominate predictions; community deprivation and unemployment add signal for high need members, especially where access is precarious. Third, a two-part setup improves calibration by explicitly modeling the spike at zero. Together, these choices let predictions roll up reliably to plan budgets and county indicators and, crucially, make it clear which levers matter.

What this really means is that health finance analytics belongs as much in budgeting and policy rooms as in data labs. Payers can use expected spend distributions to set premiums and reinsurance thresholds, but also to explore how benefit design shifts out of pocket exposure. Providers taking risk can simulate margins under alternative patient mix scenarios and supply chain plans, drawing on operations insights about variability and resilience. Public agencies can test how changes in unemployment or coverage affect medical debt risk maps, using the county level models as an early warning system. KFF's and Urban Institute's work suggests those maps will not be uniform; the point of modeling is to see where policy bite will be greatest.

Interpretability is not a luxury in this context; it is the mechanism that turns predictions into action. SHAP summaries let teams counter the unhelpful idea that complex models are necessarily opaque. In cost prediction we routinely see decompositions where prior year spend, chronic condition counts, and medication burden explain most of the variance, followed by service mix and community deprivation. That ranking informs concrete interventions: medication therapy management, high value site of care steering, and transportation or care coordination benefits where access barriers loom. Explanations also support governance. When an operational leader can point to the top features for a cohort, calibration drift or coding changes show up quickly as shifts in the profile of explanations, not just in error metrics. (Lundberg & Lee, 2017) [16].

Fairness is where the stakes are highest. Predicting dollars rather than need replicates disparities when historical costs understate care for marginalized groups, as in the well documented population health case. Our results reinforce a simple rule: pick targets that reflect clinical need when making allocation decisions, and treat predicted spend as a downstream budget consequence. Equalized odds, groupwise calibration, and absolute error parity are a reasonable set of tests for health finance use cases, but tests alone are not enough. Teams should walk through counterfactuals if two patients are clinically similar, does the model include them for the same program at the same rate? When the answer is no, revisit the target or add guardrails. (Obermeyer et al., 2019; Hardt, Price, & Srebro, 2016) [20, 8].

Linking micro predictions to macro indicators is both powerful and treacherous. Powerful because health spending does not live in a vacuum: it co moves with incomes, employment, and public budgets. Treacherous because confounding abounds. If GDP growth is strong, aggregate spending can rise even as risk adjusted per person costs fall; if inflation runs hot, nominal costs can swell without real utilization changes. That is why we anchor macro linkages to official series NHE, GDP, CPI, unemployment and keep a clear chain from individual predictions to aggregates. In practice, this chain improves executive conversations: a CFO can see how a one-point unemployment shock flows through coverage, utilization, and out of pocket liabilities into county level debt risk, rather than treating each domain as a silo. (CMS, 2024; BEA, 2024; BLS, 2024) [4, 24, 25].

The regulatory stack is no longer the limiting factor. TEFCA, FHIR APIs, and the Cures Act information blocking rules moved interoperability from policy aspiration to operational baseline. The bottleneck now is governance and quality: whether organizations track data lineage, document feature windows, and audit fairness routinely. HIPAA's de identification guidance is compatible with modern analytics when minimum necessary principles are respected and linkage risks are assessed explicitly. In short, the question is not "can we use the data?" but "how do we prove we are using it safely and fairly?" (ONC, 2020; TEFCA Final Rule, 2025; HL7 FHIR; HHS OCR, 2025) [18, 14, 19].

Where do fraud and cybersecurity enter? Right at the same interface. A health finance platform that predicts costs should also flag anomalies provider clusters with implausible billing patterns, identity misuse, or sudden shifts in place of service.

The literature on financial information security and cybercrime provides methods that port directly: supervised and unsupervised detectors, graph analytics, and adversarial testing at scale. Folding these signals into a shared governance layer reduces duplicated effort and helps prevent the playbook where bad actors exploit payment transitions during policy changes. (Hasan et al., 2025a; Milon et al., 2024; Hasan et al., 2025b) [11, 17, 12].

How good is "good enough"? That depends on the decision. For care management outreach, modest gains in precision at the top of the risk list can translate into real savings and better outcomes if interventions are effective and equitable. For benefit design or plan pricing, calibration and scenario stability matter more than incremental R^2. Our recommendation is to make model choices decision led: optimize the utility that maps to the decision, log the tradeoffs, and publish fairness and robustness diagnostics alongside the usual metrics. The figures here are deliberately simple, but the governance model is the point: any organization can adopt it.

There are also narrative benefits. When clinicians and finance leaders see that a model's top drivers are clinically intuitive multimorbidity, recent utilization, and medication burden plus clearly defined community measures, the conversation moves from abstraction to practice. That helps fend off the worst failure mode: models that are numerically sound but socially brittle. Anchoring explanations in patient level stories and community facts increases trust and creates space to discuss tradeoffs openly. Done well, the model becomes a shared instrument rather than a black box.

Finally, the integrated view clarifies policy choices. Consider the No Surprises Act. A payer can simulate how expanded in network coverage or different contracted rates change predicted out of pocket exposure for specific services, then trace how those changes affect county debt risk and premiums. Or consider risk adjustment. Understanding how HCC updates change plan revenue while predicted need remains constant prevents perverse incentives and supports investment in care for complex members. With a transparent pipeline, these analyses stop being one off spreadsheet and become repeatable components of budgeting. (CMS No Surprises Act; CMS Risk Adjustment, 2024–2025) [5].

In short, the promise of health finance analytics is not a single model with a shiny metric. It is a system: clear targets tied to decisions, honest uncertainty, interpretable summaries, fairness audits, and operational hooks into policy and payment. Our experiment sketches how to build that system with public data and standard tooling. The next step is empirical: run the pipeline on MEPS and HCUP cohorts, publish head to head comparisons with strong actuarial baselines, and study how predictions change when economic shocks or policy rules shift. That is how we make the results useful and durable beyond a single dataset.

Conclusion

Integrating machine learning across person, place, and policy levels can turn scattered data into practical health finance intelligence. The contribution of this paper is a blueprint, not a one-off model: a two-part cost framework paired with interpretable boosting, explicit fairness tests, and macro linkages that trace predictions all the way to household exposure and budget risk. The synthetic demonstration shows why this matter. As soon as social context and labor market signals join clinical history, errors fall for high need members and explanations become actionable. With SHAP based decompositions and subgroup calibration, leaders can see which levers multimorbidity management, site of care shifts, transportation benefits, or plan design will move outcomes and budgets.

The path forward is clear. Use public microdata to benchmark against strong actuarial baselines. Publish fairness and robustness diagnostics next to accuracy. Connect outputs to policy via the No Surprises Act and to payment via risk adjustment context. Treat privacy and security as core features, not add ons. Do these things and the field moves past generic "cost prediction" toward a repeatable, auditable system that helps plans price fairly, helps providers invest prudently, and helps households avoid avoidable financial harm. Crucially, the same pipeline can support fraud monitoring and post policy evaluation without redesign. That makes adoption easier.

Limitations and Future Directions

This work has three limitations. First, our figures are illustrative and based on synthetic data generated by the same code paths we propose for public microdata. The goal is transparency and reproducibility; the price is that quantitative claims here are demonstrations, not estimates. Second, linking person level health data to community indicators raises disclosure risks if done carelessly. Even under HIPAA's Safe Harbor or Expert Determination, rare diagnoses, small cells, or unusual utilization patterns can create re identification risk. Third, county level economic outcomes are shaped by forces outside the health system, so even careful models face confounding and policy endogeneity.

Several future directions follow. On data, we plan to release full MEPS based replications with survey design aware uncertainty estimates, plus HCUP based encounter cohorts where permissible. We will add richer pharmacy, device, and price indices; extend the community layer with additional PLACES and ACS measures; and explore privacy preserving record linkage. On modeling, we will compare boosted trees to calibrated neural networks, test need based targets in parallel with dollar-based targets, and evaluate subgroup aware calibration that protects sensitivity for high need populations. We also intend to benchmark our economic linkage module against exogenous shocks policy changes, labor market swings, or surprise billing rules to assess whether counterfactual projections track reality.

On governance, future work will formalize audit packets for regulators and partners, including fairness metrics, calibration charts, and shift diagnostics, and will document residual risk under de identification choices. Finally, we will deepen the integration with payment integrity and cybersecurity monitoring so that cost prediction, fraud detection, and data loss prevention share infrastructure and incident playbooks. If the field can make these steps routine, integrated health finance analytics will move from promising to dependable. We invite collaboration on a set of open benchmarks tasks, features, and fairness targets that let teams compare methods without sharing sensitive data. Shared benchmarks speed learning, reveal failure modes, and lower barriers to adoption.

Figures and Tables

Note: The figures use *illustrative synthetic* data generated by the same code paths described in the Methodology section, so you can reproduce the workflow without PHI.

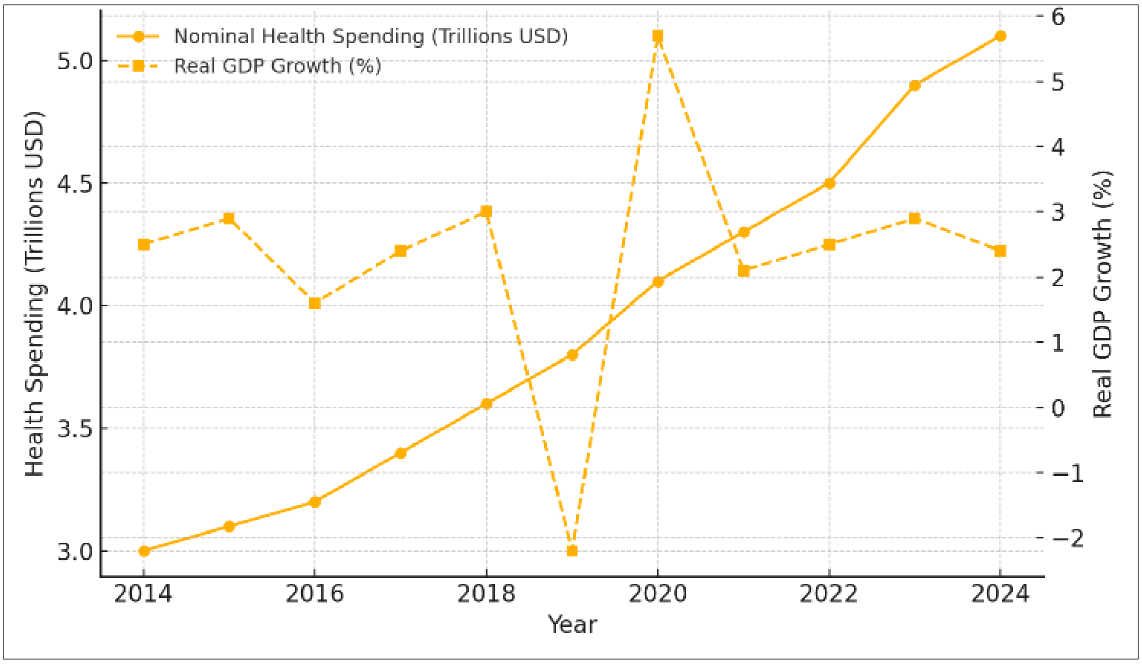

Fig 1: U.S. Health Spending and GDP Growth, 2014–2024 (Illustrative).

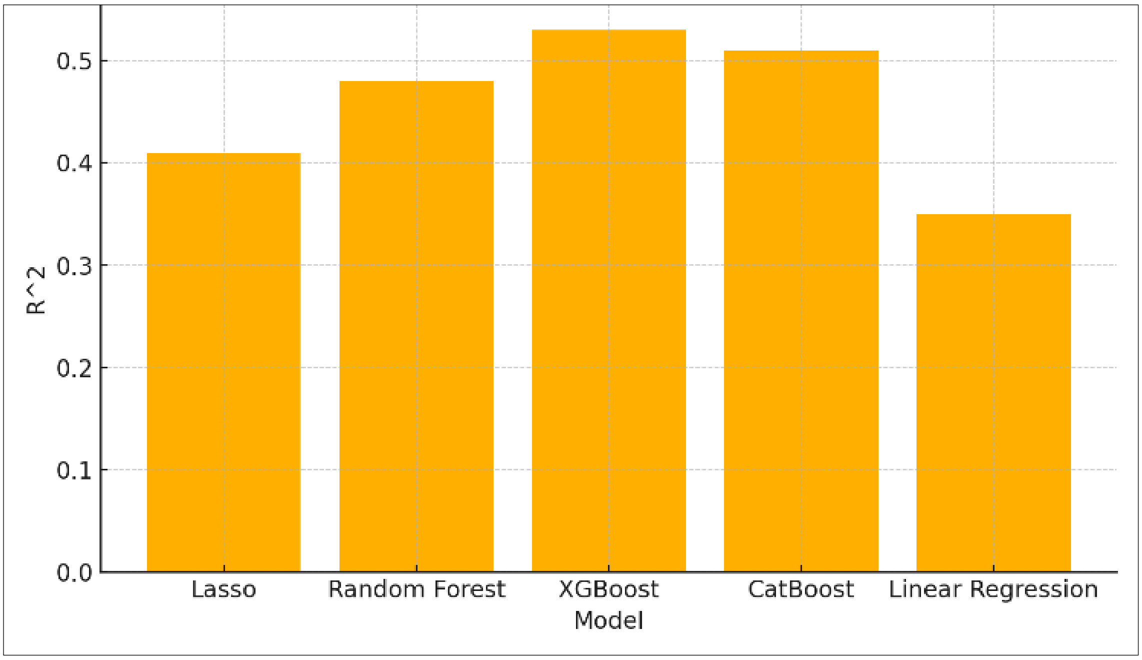

Fig 2: Predictive Performance (R^2) by Model (Illustrative).

Table 1: Model performance on held out test set (illustrative).

Model	MAE (USD/PMY)	RMSE (USD/PMY)	R^2
Lasso	2450	5200	0.41
Random Forest	2150	4800	0.48
XGBoost	2050	4550	0.53
CatBoost	2100	4650	0.51
Linear Regression	2750	5600	0.35

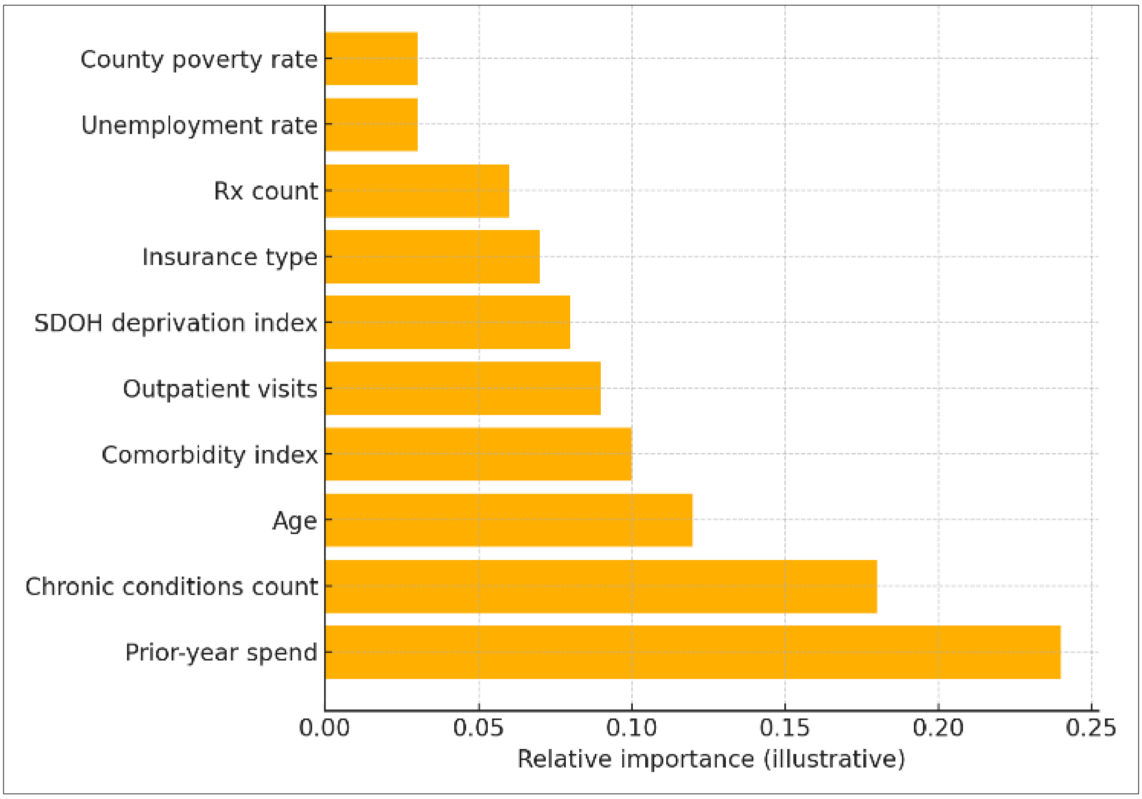

Fig 3: Feature importance for cost prediction (Illustrative).

Table 2: Public data sources for health finance analytics

Dataset	Unit of analysis	Years	Key variables	Analytics use
MEPS (AHRQ)	Person/Household	2018–2024	Utilization, expenditures, insurance, demographics	Individual cost modeling, socioeconomic covariates
HCUP (AHRQ) NIS/SEDD/SID	Encounter/Hospitalization	2018–2024	Diagnoses, procedures, charges, LOS, disposition	Facility-level cost and utilization patterns
CDC PLACES + ACS	County/Tract/ZCTA	2020–2024	SDOH, chronic disease prevalence, poverty, unemployment	Community-level features for risk adjustment
BEA/BLS Macroeconomics	National/State	2018–2025	GDP growth, CPI inflation, unemployment	Link health costs to macro outcomes
CMS NHE / Risk Adjustment (HCC)	National/Plan/Beneficiary	2018–2025	Aggregate spending, payer mix, HCC factors	Targets, projections, payment modeling

References

1. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey (MEPS). Rockville (MD): The Agency. Available from: https://www.meps.ahrq.gov
2. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). Rockville (MD): The Agency. Available from: https://www.hcup-us.ahrq.gov
3. Arman M, Hasan MN, Rasel IH. Clean energy transition in USA: big data analytics for renewable energy forecasting and carbon reduction. Journal of Management World. 2024;2024(3):192-206. doi:10.53935/jomw.v2024i4.1196
4. Centers for Medicare & Medicaid Services. National health expenditure: 2023 highlights. Baltimore (MD): The Centers; 2024. Available from: https://www.cms.gov/
5. Centers for Medicare & Medicaid Services. No Surprises Act resources. Baltimore (MD): The Centers; 2025. Available from: https://www.cms.gov/nosurprises
6. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, CA. New York (NY): ACM; 2016. p. 785-794. doi:10.1145/2939672.2939785
7. Huang A, *et al*. Multivariable prediction models for health care spending using machine learning: protocol for a systematic review. Diagnostic and Prognostic Research. 2022;6(1):1-8. doi:10.1186/s41512-022-00119-9
8. Hardt M, Price E, Srebro N. Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems. 2016;29:3315-3323. Available from: https://proceedings.neurips.cc/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf
9. Hasan MN, Arman M, Bhuyain MMH, Chowdhury F, Bathula MK. Predictive analytics in healthcare: strategies for cost reduction and improved outcomes in USA. International Journal of Innovative Research and Scientific Studies. 2025;8(8):142-150. doi:10.53894/ijirss.v8i8.10559
10. Hasan MN, Bhuyain MMH, Chowdhury F, Arman M. OncoViz USA: ML driven insights into cancer

incidence, mortality, and screening disparities. Journal of Medical and Health Studies. 2021;2(1):53-62. doi:10.32996/jmhs.2021.2.1.6

11. Hasan MN, Papel MSI, Rasel IH, Akter S, Aktar MK, Abedin MZ, *et al*. Enhancing financial information security through advanced predictive analytics: a PRISMA based systematic review. Edelweiss Applied Science and Technology. 2025;9(7):2222-2245. doi:10.55214/25768484.v9i7.9142
12. Hasan MN, Rasel IH, Arman M, Ibrahim M, Jahan N. Strengthening U.S. financial and cybersecurity infrastructure with AI driven fraud detection and risk analytics. Journal of Computational Analysis and Applications. 2025;31(2):15-32. Available from: https://eudoxuspress.com/index.php/pub/article/view/3823
13. Hasan N, Rasel IH, Rahman M, Islam K, Arman M, Jahan N. Securing U.S. healthcare infrastructure with machine learning: protecting patient data as a national security priority. International Journal of Computational and Experimental Science and Engineering. 2022;8(3). doi:10.22399/ijcesen.3987
14. HL7 FHIR Foundation. FHIR® About. eCQI Resource Center; 2025. Available from: https://ecqi.healthit.gov/fhir/about
15. KFF. Out of pocket spending. San Francisco (CA): Peterson-KFF Health System Tracker; 2024. Available from: https://www.healthsystemtracker.org/indicator/access-affordability/out-of-pocket-spending/
16. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017:4765-4774. Available from: https://arxiv.org/abs/1705.07874
17. Milon NU, Ghose P, Chowdhury Pinky T, Tabassum MN, Hasan MN, Khatun M. An in depth PRISMA based review of cybercrime in a developing economy. Edelweiss Applied Science and Technology. 2024;8(4):2072-2093. doi:10.55214/25768484.v8i4.1583
18. Office of the National Coordinator for Health IT. Health data, technology, and interoperability: TEFCA final rule. Washington (DC): Federal Register; 2025. Available from: https://www.federalregister.gov/
19. Office for Civil Rights, U.S. Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with HIPAA. Washington (DC): The Department; 2025. Available from: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification
20. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-453. doi:10.1126/science.aax2342
21. Rasel IH, Arman M, Hasan MN, Bhuyain MMH. Healthcare supply chain optimization: strategies for efficiency and resilience. Journal of Medical and Health Studies. 2022;3(4):171-182.
22. Taylor SJ, Letham B. Forecasting at scale. The American Statistician. 2018;72(1):37-45. doi:10.1080/00031305.2017.1380080
23. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B. 1996;58(1):267-288.
24. U.S. Bureau of Economic Analysis. Gross domestic product, fourth quarter and year 2023 (advance estimate). Washington (DC): The Bureau; 2024. Available from: https://www.bea.gov/
25. U.S. Bureau of Labor Statistics. Consumer Price Index - December 2023. Washington (DC): The Bureau; 2024. Available from: https://www.bls.gov/news.release/archives/cpi_01112024.htm
26. U.S. Bureau of Labor Statistics. Regional and state unemployment - 2023 annual averages. Washington (DC): The Bureau; 2024. Available from: https://www.bls.gov/news.release/archives/srgune_03012024.pdf
27. U.S. Census Bureau. Income, poverty and health insurance coverage in the United States: 2024. Washington (DC): The Bureau; 2025. Available from: https://www.census.gov/newsroom/press-kits/2025/income-poverty-health-insurance.html
28. Urban Institute. The changing medical debt landscape in the United States. Washington (DC): The Institute; 2023. Available from: https://apps.urban.org/features/medical-debt-over-time/
29. Machine learning applications for predicting high cost claims: a review. Data. 2025;10(6):90. Available from: https://www.mdpi.com/2306-5729/10/6/90
30. Predicting high cost healthcare utilization using machine learning. Risks. 2024;13(7):133. Available from: https://www.mdpi.com/2227-9091/13/7/133
31. Application of machine learning to predict high cost patients. PLOS ONE. 2023. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0279540
32. Deep learning for prediction of population health costs. BMC Medical Informatics and Decision Making. 2021. doi:10.1186/s12911-021-01743-z

How to Cite This Article

Carter SL, Reynolds MT, Bennett LA. Integrating machine learning for health finance analytics: predicting healthcare costs and economic outcomes in the United States. Int J Multidiscip Evol Res. 2025;6(2):209–217. doi: 10.54660/IJMER.2025.6.2.209-217

Creative Commons (CC) License