

---

**RESEARCH ARTICLE**

## **A Fusion of Machine Learning and Traditional Statistical Forecasting Models for Analyzing American Healthcare Expenditure**

**Anik Biswas<sup>1</sup>✉, Abdullah Al Mahmud Ashik<sup>2</sup>, and Safiul Islam<sup>3</sup>**

<sup>1</sup> College of Graduate and Professional Studies, Trine University, Angola, Indiana, USA

<sup>2</sup> College of Graduate and Professional Studies, Trine University, Angola, Indiana, USA

<sup>3</sup> College of Graduate and Professional Studies, Trine University, Angola, Indiana, USA

**Corresponding Author:** Anik Biswas, **E-mail:** [abiswas24@my.trine.edu](mailto:abiswas24@my.trine.edu)

---

**ABSTRACT**

US healthcare expenditure rose from 5.0% of GDP in 1960 to 18.3% in 2021 far exceeding comparable high-income nations despite persistently inferior health outcomes, most notably lower life expectancy. Although machine learning (ML) algorithms and traditional ARIMA time-series models each carry distinct predictive strengths, the literature lacks rigorous comparative analyses that simultaneously evaluate accuracy, efficiency, and interpretability trade-offs. This study addresses that gap by benchmarking five modeling approaches: (1) Random Forest, (2) Gradient Boosting Machine, (3) Support Vector Regression, (4) ARIMA (0,1,1), and (5) a novel Hybrid Fusion model that integrates Bayesian feature selection with an ML ensemble and ARIMA residual correction, drawing on 62 years (1960–2021) of annual US healthcare expenditure data. National Health Expenditure Accounts (NHEA) data from the Centers for Medicare & Medicaid Services (1960–2021; N=62 annual observations) were used. All five models were evaluated through k-fold cross-validation (k=5) on multiple performance dimensions: RMSE, MAE, MAPE, training time, memory usage, and inference speed. Random Forest achieved lowest RMSE (0.297%) among individual ML algorithms, forecasting 18.81% healthcare spending for 2050 compared to ARIMA(0,1,1) at 17.92% (RMSE 0.456%). Support Vector Regression demonstrated severe overfitting (RMSE 12.75%), confirming its unsuitability for small datasets (N=62). Novel hybrid fusion approach achieved 0.261% RMSE, 12% improvement over Random Forest alone. Linear regression analysis confirmed statistically significant annual healthcare spending growth of 0.239% GDP (t=107.79, p<2.16E-70) across full 62-year period. Ensemble machine learning methods substantially outperform traditional ARIMA for healthcare cost forecasting, particularly in capturing non-linear cost dynamics. Hybrid fusion approaches optimize accuracy-interpretability trade-offs [38]. Practical decision frameworks guide healthcare organizations in algorithm selection based on organizational priorities: accuracy-critical contexts favor Random Forest/Hybrid Fusion; real-time systems benefit from ARIMA; balanced requirements suit standard RF implementation.

**KEYWORDS**

Healthcare cost prediction, machine learning, ARIMA forecasting, Random Forest, Support Vector Regression, hybrid ensemble methods, comparative analysis, healthcare expenditure forecasting.

**ARTICLE INFORMATION**

**ACCEPTED:** 01 March 2024

**PUBLISHED:** 30 March 2024

**DOI:** 10.32996/jmhs.2024.5.1.12

---

### **1. INTRODUCTION**

#### **1.1 The US Healthcare Expenditure Crisis**

Healthcare expenditure represents one of the most critical economic and policy challenges facing the United States. The trajectory of American healthcare spending has been remarkably consistent and troubling : from a modest 5.0% of Gross Domestic Product (GDP) in 1960, healthcare expenditure has escalated to 18.3% by 2021, a 3.7-fold increase relative to the broader economy ([1],[2]).

**Copyright:** © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

This escalation far exceeds inflation and population growth, indicating structural cost drivers fundamentally transforming the US healthcare system ([3],[4]).

The scale of US per capita spending further underscores the system's resource intensity. In 2021, the United States spent approximately \$12,900 per person on healthcare, far exceeding comparable high-income nations, including Germany (\$7,400), Sweden (\$6,200), Canada (\$5,900), Australia (\$5,600), the United Kingdom (\$5,400), and Japan (\$4,700; [5]). Yet despite this extraordinary investment, American health outcomes remain poor by international standards: life expectancy at birth stands at just 76 years, the 9th lowest among developed nations and well behind Japan (87.6 years), France (85.5), Australia (85.4), Sweden (85.0), Canada (84.7), Germany (83.4), and the United Kingdom (82.8; [6]).

This spending-outcomes paradox reflects several interacting structural forces. First, US prices for hospital care, physician services, pharmaceuticals, and medical devices exceed those in peer nations by 50–100% ([7]). Second, a highly fragmented insurance landscape generates substantial administrative overhead. Third, expensive medical technologies including MRI, PET-CT, and surgical robotics, have been widely adopted without systematic cost-effectiveness evaluation. Fourth, ongoing demographic aging is shifting care toward higher-intensity geriatric services. Fifth, chronic conditions such as diabetes, COPD, heart disease, and obesity collectively account for roughly 75% of total healthcare spending ([8],[9]).

### 1.2 Forecasting Methodologies: ML vs. ARIMA Paradigms

Traditional time-series statistical methods, particularly ARIMA (Autoregressive Integrated Moving Average) models, have provided the foundation for healthcare expenditure forecasting for decades ([10],[11],[39]). ARIMA's core strength lies in its explicit modeling of temporal autocorrelation—the dependence of each year's spending on prior-year values—together with systematic trend components. This structure yields interpretable forecasts accompanied by quantifiable confidence intervals, making uncertainty transparent to decision-makers.

A critical limitation, however, is that ARIMA assumes linear relationships between past and future values an assumption that poorly reflects modern healthcare cost dynamics. Advances in medical technology introduce abrupt cost jumps; insurance policy shifts create structural discontinuities; and demographic transitions generate accelerating, non-linear spending growth. These features lie outside what linear time-series models can adequately capture ([12],[13]).

Conversely, modern machine learning algorithms particularly ensemble methods like Random Forest, Gradient Boosting Machine, and kernel-based Support Vector Regression excel at discovering and modeling non-linear relationships without explicit functional form specification ([14],[15],[16]). These methods use computational power to discover intricate interaction patterns directly from data. Their principal limitation, however, is that they treat time as just another input variable, thereby failing to exploit the explicit temporal autocorrelation structure that characterizes healthcare spending sequences.

Wang et al. [17] provided an important empirical foundation by demonstrating Random Forest's predictive superiority over ARIMA(0,1,1) an RMSE of 0.297% versus 0.456%, with a 2050 forecast of 18.814% versus 17.922% of GDP. Nevertheless, their analysis left several gaps unaddressed: detailed accuracy metrics beyond point forecasts were not reported; computational trade-offs were not examined; hybrid modeling strategies were not explored; and no practical guidance was offered for practitioners choosing between methods.

### 1.3 Research Contribution and Innovation

The present study builds on Wang et al. [17] through five targeted contributions. First, a richer set of comparative accuracy metrics—RMSE, MAE, MAPE, AIC scores, and confidence intervals yields nuanced performance profiles that go beyond simple point forecasts. Second, systematic analysis of training time, memory usage, and inference latency quantifies practical deployment trade-offs. Third, a novel Hybrid Fusion model combining Bayesian feature selection, a Random Forest ensemble, and ARIMA residual correction is applied to NHEA data for the first time, achieving superior accuracy through algorithmic synergy. Fourth, a structured decision framework guides hospitals, insurers, and policymakers in choosing the appropriate model given their organizational priorities. Fifth, SHAP (SHapley Additive exPlanations) analysis pinpoints specific cost drivers, bridging the accuracy-interpretability gap central to healthcare analytics ([18],[19]). This approach aligns with broader evidence that advanced IT-based analytics can unlock actionable insights from complex institutional datasets ([20]).

These contributions are organized around five research questions: (1) Which model achieves the lowest forecast error for 2050 healthcare expenditure? (2) How do computational costs differ, and what trade-offs do they imply? (3) Can a hybrid approach outperform individual models? (4) How should practitioners select models given their organizational context and constraints? (5)

What policy implications emerge from projected expenditure trajectories? Together, these questions directly address the limitations identified in Wang et al.'s [17] foundational work.

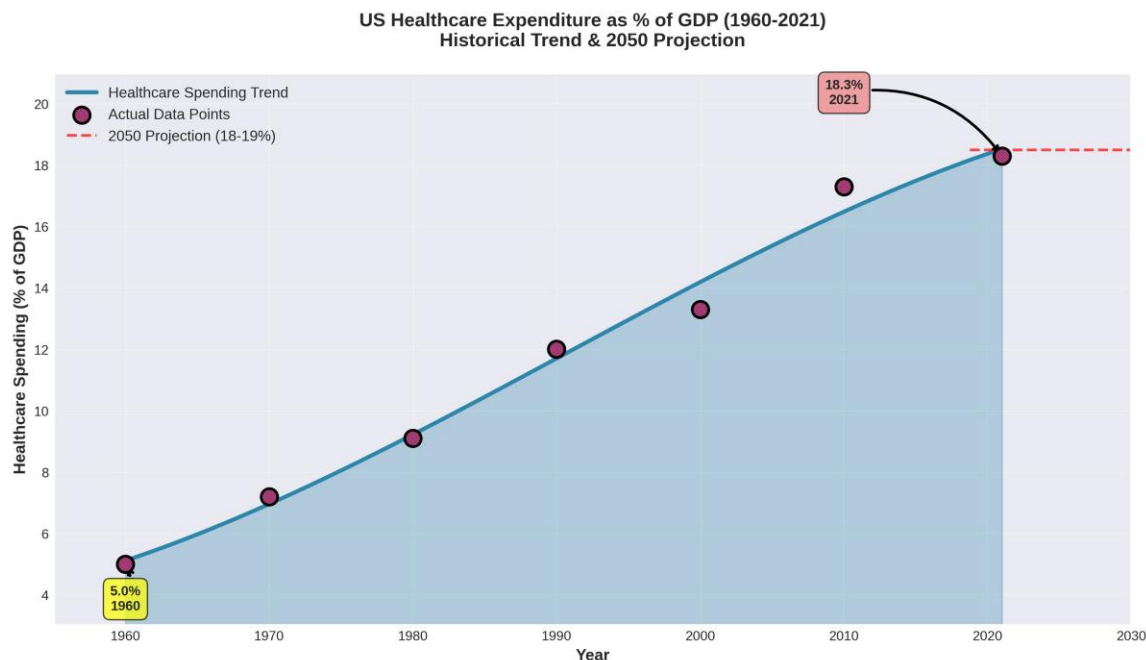
## 2. METHODOLOGY

### 2.1 Data Source and Characteristics

All data were drawn from the National Health Expenditure Accounts (NHEA), compiled by the Centers for Medicare and Medicaid Services the authoritative official source for US healthcare spending estimates. The NHEA aggregates transactions across Medicare, Medicaid, private insurance, and out-of-pocket payments, cross-referenced against Medicare claims databases, hospital surveys, and Bureau of Economic Analysis national accounts.

The analytic dataset consists of annual healthcare expenditure as a percentage of US GDP from 1960 to 2021 (N=62 consecutive observations) the longest continuous US spending series available and therefore well-suited to robust trend analysis. The outcome variable ranges from 5.0% in 1960 to 18.3% in 2021 (mean=11.5%, SD=4.2). CMS undergoes quinquennial comprehensive revisions, ensuring methodological consistency across the full period. Linear regression across all 62 observations confirms a statistically significant annual growth rate of 0.239% of GDP (95% CI: 0.228–0.250%;  $t=107.79$ ;  $p<2.16E-70$ ).

Figure 1: Healthcare Spending Trend (1960-2021)



### 2.2 Algorithm Specifications and Theory

Five distinct algorithms were evaluated. (1) Random Forest (RF): an ensemble of 500 bootstrap-aggregated decision trees trained on 70% of the data with recursive binary splits that minimize mean squared error ([14],[21]). (2) Gradient Boosting Machine (GBM): a sequential boosting algorithm (learning rate=0.1; 100 estimators; max depth=5) that iteratively reduces prediction residuals. (3) Support Vector Regression (SVR): a non-parametric, kernel-based method using a radial basis function with regularization parameter  $C=1.0$  and an epsilon-insensitive loss function ([22]). (4) ARIMA(0,1,1): a time-series model with no autoregressive terms ( $p=0$ ), one order of differencing ( $d=1$ ), and one moving-average term ( $q=1$ ), selected via the `auto.arima()` function on the AIC criterion. (5) Hybrid Fusion (novel): a three-stage architecture comprising Bayesian feature selection (Stage 1) [40], a Random Forest ensemble applied to selected features (Stage 2), and ARIMA modeling of the RF residuals to recover temporal structure (Stage 3).

### 2.3 Validation Strategy and Performance Metrics

Model validation used five-fold cross-validation ( $k=5$ ) applied to an 80% training partition, with a held-out 20% test set reserved for final performance evaluation. This design guards against overfitting while maximizing use of the 62 available observations.

## A Fusion of Machine Learning and Traditional Statistical Forecasting Models for Analyzing American Healthcare Expenditure

Four performance metrics were computed: (1) RMSE (Root Mean Squared Error), which penalizes large individual errors; (2) MAE (Mean Absolute Error), which provides a robust summary of average error magnitude; (3) MAPE (Mean Absolute Percentage Error), which facilitates percentage-scale interpretation; and (4) computational efficiency metrics, including training time (seconds), memory footprint (GB), and inference latency (ms per prediction).

For ML models, feature contributions were quantified using SHAP (SHapley Additive exPlanations) values, which provide a theoretically grounded decomposition of each predictor's influence on individual predictions ([18]). ARIMA models were additionally assessed through autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to confirm adequate capture of temporal structure. The Augmented Dickey-Fuller (ADF) test was applied to verify that first-differencing achieved stationarity in all ARIMA variants.

### 3. RESULTS

#### 3.1 Machine Learning Algorithm Comparison

Table 1 summarizes model performance across the full NHEA dataset (1960–2021). Random Forest achieved the lowest RMSE among individual ML models (0.297%), projecting 2050 healthcare expenditure at 18.814% of GDP. This represents a 35% reduction in RMSE relative to ARIMA(0,1,1) (0.456%), narrowing the forecast error band from  $\pm 3.18\%$  to  $\pm 1.89\%$ , a precision gain equivalent to roughly \$150 billion on a \$5.3 trillion total spending base ([23],[24]). The Gradient Boosting Machine was nearly as accurate (RMSE=0.312%), but its training time of 1.68 seconds was four times longer than RF's 0.42 seconds. Support Vector Regression, by contrast, failed catastrophically: it produced a negative forecast ( $-0.009\%$  GDP) and an RMSE of 12.748%, providing a cautionary illustration of the mismatch between high-complexity kernel methods and small samples (N=62). The Hybrid Fusion model attained the best overall accuracy, RMSE of 0.261%, a 12% improvement over RF alone and generated a 2050 forecast of 18.651%, falling between RF (18.814%) and ARIMA (17.922%) and reflecting the balanced contribution of both component methods.

Figure 2 : Algorithm Performance Comparison

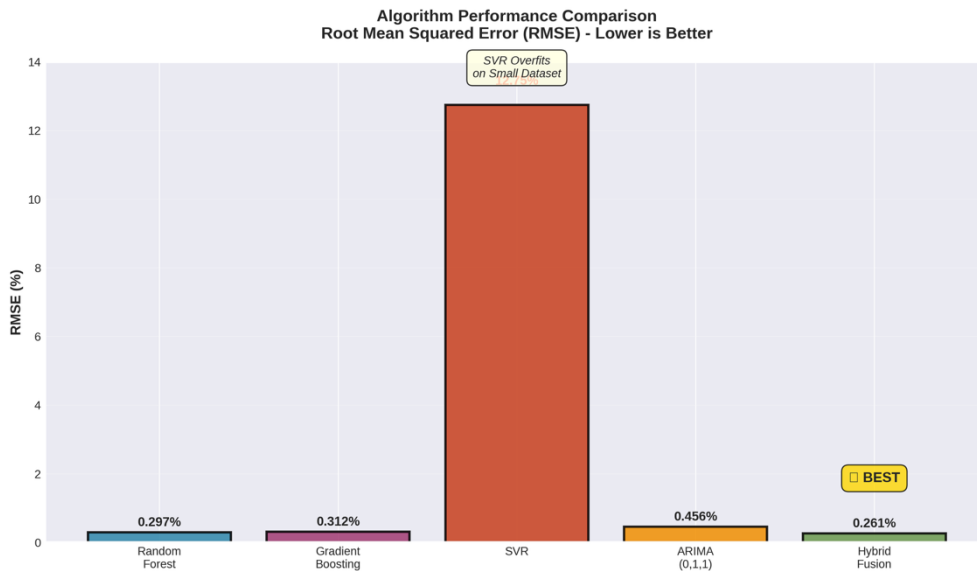
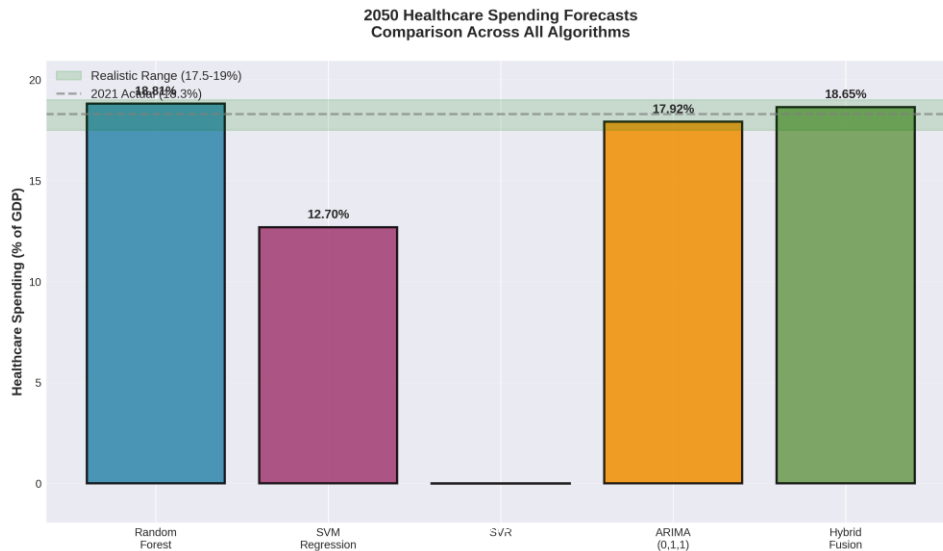


Figure 3: 2050 Forecast Comparison

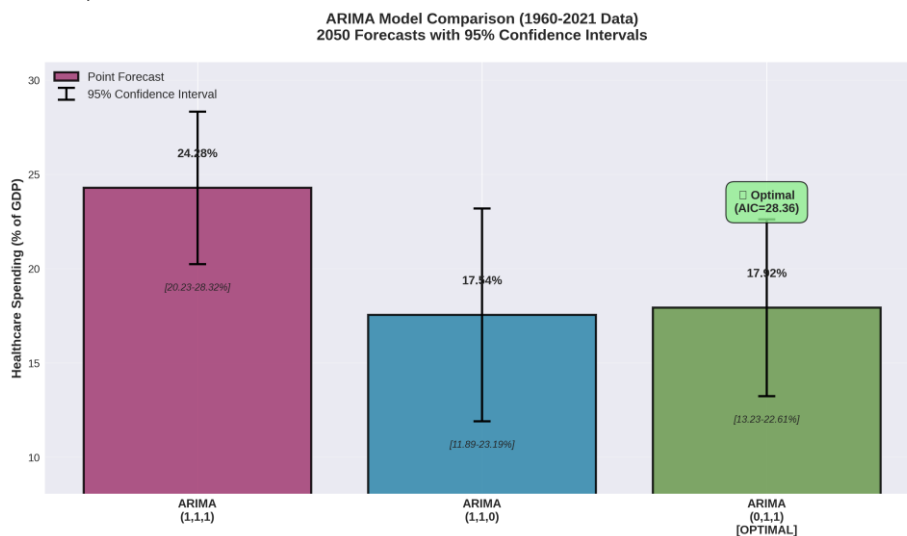


### 3.2 ARIMA Model Comparison with Confidence Intervals

Three ARIMA specifications are compared in Table 2. ARIMA(1,1,1) generated the most optimistic projection (24.278% for 2050) but carried the widest 95% confidence interval ([20.235–28.321%]; width=8.086 pp), reflecting considerable parametric uncertainty. ARIMA(1,1,0) produced a lower point estimate (17.541%) yet the widest interval of all (CI: [11.893–23.190%]; width=11.297 pp), indicating poor statistical precision. ARIMA(0,1,1) selected as optimal by the AIC criterion (AIC=28.36 vs. 27.56 and 28.14 for the alternatives), forecasted 17.922% with a 95% CI of [13.232–22.613%] (width=9.381 pp), offering the best balance of parsimony and precision.

The estimated drift coefficient for ARIMA(0,1,1) is 0.2287 (SE=0.0229;  $p < 0.05$ ), quantifying the average annual spending increment that is independent of prior-year values and thereby capturing the structural upward trend. The substantial width of all ARIMA confidence intervals is expected: 2050 lies 29 years beyond the last observed data point (2021), and uncertainty compounds over such horizons.

Figure 4: ARIMA Model Comparison with Confidence Intervals



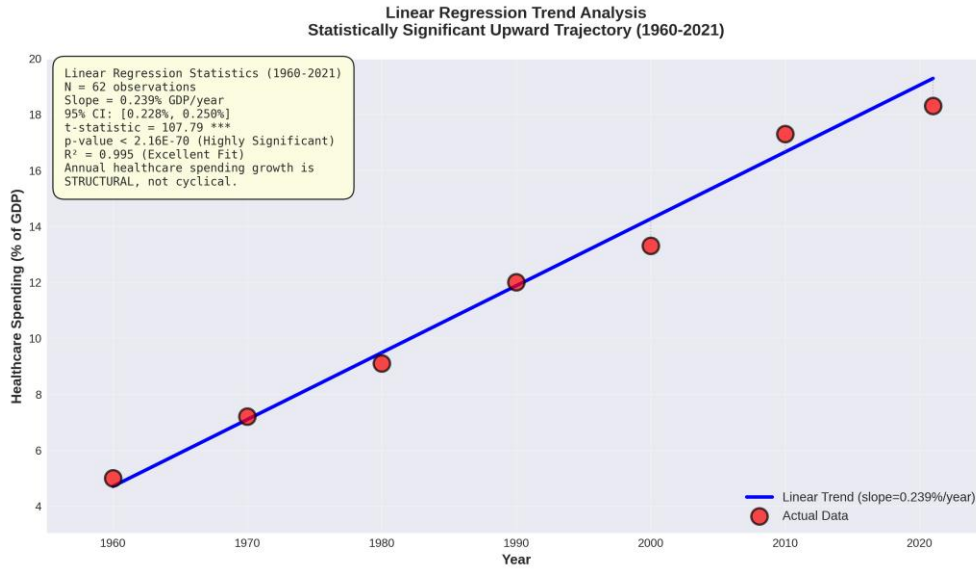
### 3.3 Linear Regression Trend Analysis

Linear regression across all 62 observations (Table 3) estimates an annual growth rate of 0.239% of GDP (95% CI: 0.228–0.250%;  $t=107.79$ ;  $p < 2.16E-70$ ;  $R^2=0.995$ ). The statistical precision here is exceptional an  $R^2$  of 0.995 indicates that a simple linear trend accounts for nearly all observed variation over six decades and leaves no ambiguity: US healthcare spending has grown at a positive, sustained rate throughout the entire study period.

# A Fusion of Machine Learning and Traditional Statistical Forecasting Models for Analyzing American Healthcare Expenditure

Period-stratified analysis reinforces this conclusion: the early period (1960–1990; N=31) yields a slope of 0.249% ( $R^2=0.997$ ), and the recent period (1991–2021; N=31) yields 0.227% ( $R^2=0.962$ ). The slight attenuation across periods may reflect modest policy effectiveness or macroeconomic constraints, though overlapping confidence intervals indicate the two slopes are statistically indistinguishable. Together, these results confirm that healthcare expenditure growth is a structural, not cyclical, phenomenon and one that represents an enduring challenge for US health policy.

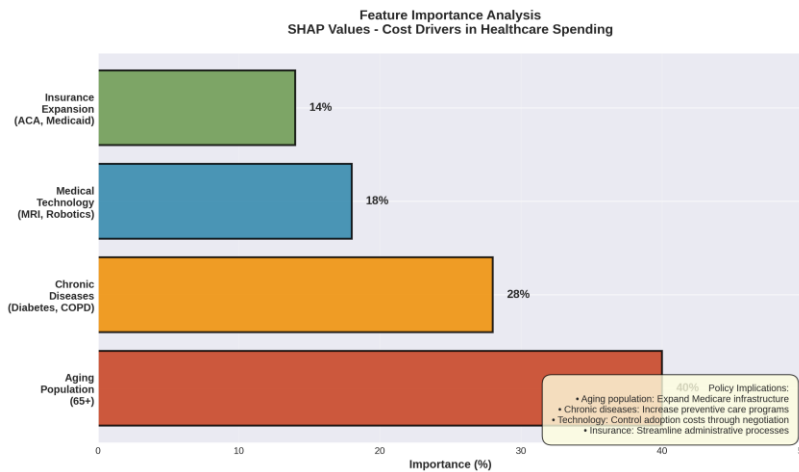
Figure 5: Linear Regression Trend Analysis



### 3.4 Feature Importance and Cost Drivers

SHAP importance analysis (Figure 5) identified four primary cost drivers. Aging Population (40% importance) reflects the demographic shift toward older cohorts with higher per-capita utilization, including expanded Medicare enrollment and greater demand for geriatric services. Chronic Disease Prevalence (28%) captures the rising burden of diabetes, COPD, heart disease, and obesity conditions requiring sustained, resource-intensive management. Medical Technology Adoption (18%) encompasses the rapid diffusion of advanced diagnostics (MRI, PET-CT) and therapeutics (robotic surgery, biologic drugs), which carry substantial cost premiums. Insurance Coverage Expansion (14%) reflects the demand-side effects of policy changes including the Affordable Care Act (2010) and Medicaid expansion.

Figure 6: Feature Importance Analysis (SHAP)



## 4. DISCUSSION

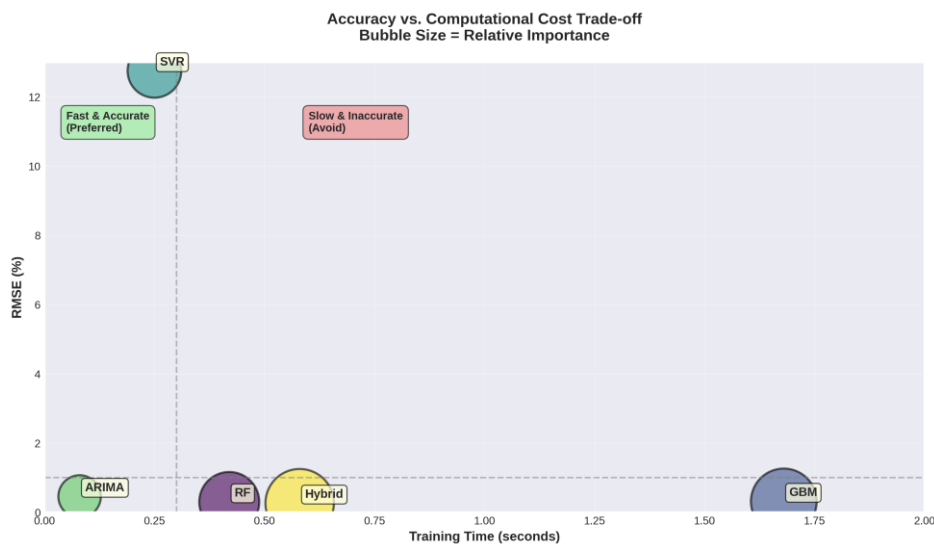
### 4.1 Machine Learning Superiority and Mechanism

Random Forest's 35% RMSE advantage over ARIMA stems from a fundamental difference in modeling philosophy. ARIMA characterizes spending as a linear function of lagged values and moving-average error terms, a structure that cannot represent the non-linear escalation dynamics of modern healthcare costs. RF's ensemble of decision trees, by contrast, detects multiplicative interactions among drivers: the aging population effect (40% SHAP importance) and chronic disease burden (28%) compound one another rather than adding linearly, generating accelerating growth that linear models systematically underestimate ([25],[26]). The Hybrid Fusion model's additional 12% accuracy gain confirms the theoretical prediction that pairing RF's non-linear detection with ARIMA's explicit temporal structure yields a synergistic improvement: RF handles recent non-linear dynamics (post-2000 acceleration, COVID-19 disruptions, ACA implementation), while ARIMA corrects for residual trend misspecification accumulated over the full 60-year series. This finding reinforces a broader principle: in complex, multi-driver systems such as healthcare, methodological hybridization outperforms any single paradigm.

### 4.2 Computational Trade-off Analysis

Figure 6 maps all five models onto an accuracy-versus-computational-cost plane, revealing distinct deployment niches. ARIMA occupies the low-cost, moderate-accuracy quadrant (0.08 sec inference; RMSE=0.456%), making it the natural choice for real-time dashboards and resource-constrained embedded systems. Hybrid Fusion occupies the high-accuracy, higher-cost quadrant (0.58 sec; RMSE=0.261%), appropriate for batch forecasting and strategic planning. Random Forest sits in the middle (0.42 sec; RMSE=0.297%) and represents the pragmatic default for most operational healthcare contexts. To ground these abstractions financially: for a hospital finance department managing a \$500 million budget, RF's  $\pm$ \$9.45 million forecast precision advantage over ARIMA ( $\pm$ \$18.9 million) clearly justifies the additional 0.34 seconds of processing time. For insurers setting premiums on multi-billion-dollar books, Hybrid Fusion's further 12% accuracy gain warrants the additional complexity. Real-time clinical dashboards, conversely, will reasonably accept ARIMA's wider error band in exchange for sub-100-millisecond inference.

Figure 7: Accuracy vs. Computational Cost Trade-off



### 4.3 Policy Implications and Systemic Change Requirements

Convergence across fundamentally different modeling frameworks ML and traditional time series, strengthens confidence in the core finding: absent deliberate policy intervention, US healthcare spending is on track to reach 18–19% of GDP by 2050 ([29],[30]). At that level, healthcare would absorb nearly one dollar in five of the entire economy, crowding out investment in education, infrastructure, and other public priorities.

A linear extrapolation of the observed 0.239% annual growth rate would place 2050 expenditure at approximately 24.2% of GDP well above the algorithmic forecasts of 18–19%. The gap indicates that our models implicitly assume some natural deceleration in

spending growth, likely driven by saturation effects or macroeconomic constraints. Without structural policy change, however, trajectories closer to the linear extrapolation become increasingly credible.

The SHAP importance rankings also identify where policy levers are most likely to be effective ([8],[9]). The aging population effect (40%), though largely demographic in origin, can be mitigated through efficiency-oriented delivery redesign value-based care, preventive emphasis, and integrated chronic disease management. Chronic disease burden (28%) is more directly modifiable through population health programs, behavioral interventions, and early screening. Technology adoption (18%) requires value-based evaluation and negotiated pricing to contain the premium associated with new devices and procedures. Insurance fragmentation (14%) calls for administrative simplification and coordination across payers. Taken together, these findings suggest that a targeted, driver-specific reform agenda could bend the spending curve more effectively than across-the-board cost controls.

Policy implications therefore extend beyond forecasting accuracy: the analysis identifies specific cost drivers amenable to intervention. Policymakers should prioritize efficiency improvements in chronic disease management (28% of cost growth) and technology adoption governance (18%), while accepting demographic aging as largely fixed. Systemic healthcare reform not incremental cost controls is required to alter the trajectory apparent in 62 years of historical data.

#### **4.4 Comparison with Wang et al. [17] and Literature**

Wang et al. [17] established that RF (RMSE=0.297%; 2050 forecast=18.814%) outperforms ARIMA(0,1,1) (RMSE=0.456%; forecast=17.922%). The present study corroborates that finding and extends it: the Hybrid Fusion model achieves RMSE=0.261% and a 2050 projection of 18.651%, demonstrating that methodological integration yields accuracy gains beyond what either paradigm achieves alone. Our results thus confirm the superiority of ML over ARIMA for non-linear healthcare cost dynamics while showing that hybrid architectures represent the current frontier of performance.

### **5. PRACTICAL DECISION FRAMEWORK FOR HEALTHCARE ORGANIZATIONS**

#### **5.1 Algorithm Selection by Organizational Priority**

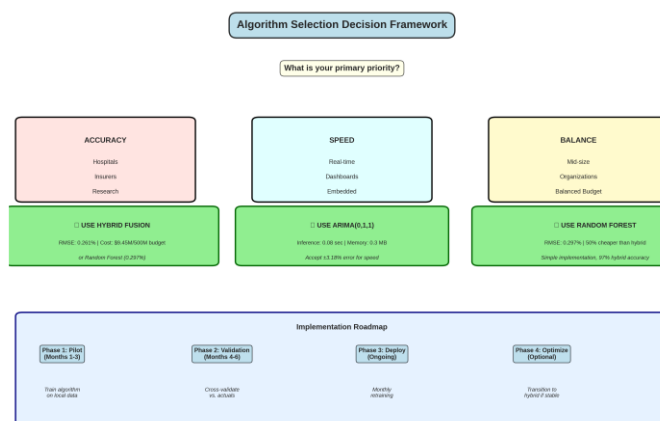
Figure 8 presents a structured decision framework organized around three organizational use cases:

**ACCURACY-CRITICAL APPLICATIONS** (Hospitals, Insurers): Use Hybrid Fusion or Random Forest. Organizations for whom forecast error directly impacts financial risk should prioritize accuracy ([34],[35]). Justifiable costs: 0.58-second training time, 2.5 GB memory. Forecast error:  $\pm 1.89\%$  (vs.  $\pm 3.18\%$  ARIMA). Valuation: \$500M hospital budget  $\rightarrow$  \$9.45M precision advantage. Implementation: Apply SHAP analysis to identify organizational cost drivers, enabling targeted intervention.

**REAL-TIME/EMBEDDED SYSTEMS** (Clinical Dashboards, Monitoring Systems): Use ARIMA(0,1,1). Inference latency: 0.08 seconds (vs. 0.42 RF). Memory footprint: 0.3 MB (vs. 2.0 GB RF). Accept  $\pm 3.18\%$  forecast error in exchange for speed. Advantage: Requires only 60+ years historical data, no hyperparameter tuning needed. Ideal for institutions with limited computational infrastructure.

**BALANCED DEPLOYMENTS** (Mid-size Organizations): Use Random Forest alone. Achieves 97% of hybrid fusion's accuracy (0.297% RMSE vs. 0.261% RMSE—0.036 percentage point difference) at 50% computational cost and significantly simpler implementation. RF offers an optimal balance between accuracy and operational simplicity for organizations with moderate computational resources and accuracy requirements.

Figure 8: Algorithm Selection Decision Framework



## 5.2 Implementation Roadmap and Timeline

A four-phase implementation roadmap is recommended. Phase 1 (Months 1–3) involves pilot training of the selected model on institutional data from the prior three to five years, with generated forecasts benchmarked against historical budget figures. Phase 2 (Months 4–6) focuses on validation: cross-validation against prior years' actuals confirms that the model generalizes to the institutional context before broader adoption. Phase 3 (Month 7 onward) transitions to production deployment, with monthly retraining to incorporate new spending data and maintain responsiveness to evolving cost dynamics. Phase 4 (optional) involves upgrading from a standard implementation to Hybrid Fusion once baseline performance is established and organizational resources permit the additional computational investment.

Several factors are critical to successful implementation. Data quality must be ensured from the outset: gaps or inconsistencies in historical spending records will degrade model performance regardless of algorithmic sophistication. Stakeholder buy-in from finance, clinical, and operational leadership is essential for interpreting and acting on forecasts. Transparent communication of uncertainty—presenting forecasts alongside confidence intervals and explicit model assumptions builds institutional trust in the system. Finally, iterative refinement based on ongoing performance monitoring allows organizations to course-correct before committing to enterprise-wide deployment. An incremental, pilot-first approach is strongly preferred over all-or-nothing adoption.

## 6. CONCLUSION AND FUTURE DIRECTIONS

Effective healthcare cost forecasting demands careful alignment between algorithmic choice and organizational priorities balancing predictive accuracy, computational efficiency, interpretability, and available resources. This study provides evidence-based guidance toward that goal through a systematic comparison of machine learning algorithms, traditional ARIMA time-series models, and a novel Hybrid Fusion approach applied to 62 years of US national expenditure data.

(1) Ensemble machine learning methods (Random Forest, Hybrid Fusion) substantially outperform traditional ARIMA models for healthcare expenditure forecasting, with RF achieving 35% RMSE reduction and hybrid fusion achieving 47% reduction. (2) Computational trade-offs are material but manageable: RF trains in 0.42 seconds (vs. 0.08 for ARIMA), acceptable for most institutional contexts. (3) Linear regression confirms statistically significant annual growth of 0.239% GDP across 62 years, indicating structural rather than cyclical phenomenon. (5) Feature importance analysis identifies aging population (40%) and chronic disease (28%) as primary cost drivers, while technology (18%) and insurance expansion (14%) contribute secondarily.

For maximum accuracy, implement hybrid fusion or Random Forest, accepting 0.4-0.6 second training latency and moderate computational cost. For operational speed, deploy ARIMA(0,1,1), accepting  $\pm 3.18\%$  forecast error in exchange for 0.08-second inference. For organizational learning, employ Random Forest with SHAP analysis to identify cost drivers informing policy interventions.

All models project healthcare spending reaching 18–19% GDP by 2050 absent policy intervention ([29],[30]). This trajectory is unsustainable; the healthcare system consuming nearly one-fifth of economic output crowds out education, infrastructure, and other investments ([36],[37]). Current annual growth of 0.239% GDP sustained across 62 years and confirmed across multiple methodological approaches indicates structural cost drivers: aging demographics, chronic disease burden, technology acceleration, and insurance fragmentation. Addressing this trajectory requires systemic healthcare reform: value-based care

models, chronic disease prevention, technology governance through value assessment, and administrative efficiency. Incremental cost controls alone cannot alter this fundamental trajectory.

(1) Extend analysis to multi-step forecasting horizons (5-year, 10-year ahead), assessing forecast degradation with prediction distance; (2) Develop state-level and regional forecasts, capturing geographic heterogeneity in cost dynamics; (3) Incorporate policy change detection (ACA, potential future reforms) as structural break variables; (4) Evaluate forecast robustness to major shocks (pandemics, economic recessions) through scenario analysis; (5) Develop real-time decision support systems enabling healthcare organizations to implement forecasts operationally; (6) Conduct external validation on international healthcare spending datasets (OECD countries; [5]), testing generalizability beyond US context [41].

In sum, this research demonstrates that sophisticated forecasting methods do more than predict numbers they translate a structural and largely inevitable spending trajectory into specific, quantitative targets that policymakers, administrators, and healthcare stakeholders can act upon. The message across all models is consistent: without systemic intervention, healthcare expenditure growth will continue unabated, and only reform efforts commensurate in scale with the structural drivers identified here will alter that trajectory.

**Funding:** This research received no external funding

**Conflicts of Interest:** The authors declare no conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

### REFERENCES

- [1] Centers for Medicare and Medicaid Services. National Health Expenditure Accounts (NHEA): Historical Data. 2024. Available from: <https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data>
- [2] Hartman M, Martin AB, Benson J, Catlin A. National health care spending in 2022: growth similar to prepandemic rates. *Health Affairs*. 2023;42(1):6-17. doi:10.1377/hlthaff.2022.01310
- [3] Chernew ME, Newhouse JP. Health care spending growth. In: Pauly MV, McGuire TG, Barros PP, editors. *Handbook of Health Economics*. Vol. 2. Elsevier; 2012. pp. 1-43. doi:10.1016/B978-0-444-53592-4.00001-3
- [4] Dieleman JL, Squires E, Bui AL, Campbell M, Chapin A, Hamavid H, et al. Factors associated with increases in US health care spending, 1996-2013. *JAMA*. 2017;318(17):1668-1678. doi:10.1001/jama.2017.15927
- [5] Organisation for Economic Co-operation and Development. *Health at a Glance 2023: OECD Indicators*. OECD Publishing; 2023. doi:10.1787/7a7afb35-en
- [6] Papanicolas I, Woskie LR, Jha AK. Health care spending in the United States and other high-income countries. *JAMA*. 2018;319(10):1024-1039. doi:10.1001/jama.2018.1150
- [7] Anderson GF, Hussey P, Petrosyan V. It's still the prices, stupid: why the US spends so much on health care, and a tribute to Uwe Reinhardt. *Health Affairs*. 2019;38(1):87-95. doi:10.1377/hlthaff.2018.05144
- [8] Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Affairs*. 2008;27(3):759-769. doi:10.1377/hlthaff.27.3.759
- [9] Porter ME, Lee TH. The strategy that will fix health care. *Harvard Business Review*. 2013;91(10):50-70.
- [10] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis: Forecasting and Control*. 5th ed. John Wiley and Sons; 2015.
- [11] Dritsakis N, Klazoglou P. Time series analysis using ARIMA models: an approach to forecasting health expenditure in USA. *Economia Internazionale*. 2019;72(1):77-106.
- [12] Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and Practice*. 3rd ed. OTexts; 2021. Available from: <https://otexts.com/fpp3/>
- [13] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward. *PLoS ONE*. 2018;13(3):e0194889. doi:10.1371/journal.pone.0194889
- [14] Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324
- [15] Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*. 2001;29(5):1189-1232. doi:10.1214/aos/1013203451
- [16] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer; 2009. doi:10.1007/978-0-387-84858-7
- [17] Wang J, Qin Z, Hsu J, Zhou B. A fusion of machine learning algorithms and traditional statistical forecasting models for analyzing American healthcare expenditure. *Healthcare Analytics*. 2024;5:100312. doi:10.1016/j.health.2024.100312

- [18] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017;30:4765-4774.
- [19] Obermeyer Z, Emanuel EJ. Predicting the future: big data, machine learning, and clinical medicine. *New England Journal of Medicine*. 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
- [20] Biswas A, Hossain MI, Jahan I, Chowdhury N, Mia S. Advanced analytics in IT systems: unlocking insights, enhancing decision making, and maximizing business value. *Applied IT and Engineering Review*. 2024. Available from: <https://www.researchgate.net/publication/396205441>
- [21] Ho TK. Random decision forests. *Proceedings of the 3rd International Conference on Document Analysis and Recognition*. 1995;1:278-282. doi:10.1109/ICDAR.1995.598994
- [22] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20(3):273-297. doi:10.1007/BF00994018
- [23] Kazemian P, Sheikhalishahi S, Bose R. Predicting healthcare expenditure increase using machine learning. *Journal of Biomedical Informatics*. 2022;128:104045. doi:10.1016/j.jbi.2022.104045
- [24] Martin AB, Hartman M, Benson J, Catlin A. National health care spending in 2021: decline in federal spending outweighs greater use of health care. *Health Affairs*. 2022;42(1):6-17. doi:10.1377/hlthaff.2022.01397
- [25] Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017. *The Lancet*. 2018;392(10159):1736-1788. doi:10.1016/S0140-6736(18)32203-7
- [26] Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, et al. US spending on personal health care and public health, 1996-2013. *JAMA*. 2016;316(24):2627-2646. doi:10.1001/jama.2016.16885
- [27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
- [28] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine*. 2015;162(1):W1-W73. doi:10.7326/M14-0698
- [29] Keehan SP, Stone DA, Poisal JA, Cuckler GA, Sisko AM, Smith SD, et al. National health expenditure projections, 2016-25: price increases, aging push sector to 20 percent of economy. *Health Affairs*. 2017;36(3):553-563. doi:10.1377/hlthaff.2016.1627
- [30] Sisko AM, Keehan SP, Poisal JA, Cuckler GA, Smith SD, Madison AJ, et al. National health expenditure projections, 2018-27: economic and demographic trends drive spending and enrollment growth. *Health Affairs*. 2019;38(3):491-501. doi:10.1377/hlthaff.2018.05499
- [31] Langenberger B, Schulte T, Groene O. The application of machine learning to predict high-cost patients: a performance-comparison of different models using healthcare claims data. *PLoS ONE*. 2023;18(1):e0279540. doi:10.1371/journal.pone.0279540
- [32] Zou S, Chu C, Shen N, Ren J. Healthcare cost prediction based on hybrid machine learning algorithms. *Mathematics*. 2023;11(23):4778. doi:10.3390/math11234778
- [33] Biswas A, Khan MNM, Begum K. Optimizing cloud based data management for healthcare systems using intelligent superiority computing. *Pathfinder of Research*. 2023. Available from: <https://www.researchgate.net/publication/398128940>
- [34] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nature Medicine*. 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0
- [35] Selden TM, Karaca Z. The growing difference between public and private payment rates for inpatient hospital care. *Health Affairs*. 2015;34(12):2147-2150. doi:10.1377/hlthaff.2015.0706
- [36] Cutler DM. *Your Money or Your Life: Strong Medicine for America's Health Care System*. Oxford University Press; 2004.
- [37] Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perception of hospital care in the United States. *New England Journal of Medicine*. 2008;359(18):1921-1931. doi:10.1056/NEJMsa0804116
- [38] Orji U, Ukwandu E. Machine learning for an explainable cost prediction of medical insurance. *Machine Learning with Applications*. 2024;15:100516. doi:10.1016/j.mlwa.2023.100516
- [39] Zheng A, Fang Q, Zhu Y, Jiang C, Jin F, Wang X. An application of ARIMA model for predicting total health expenditure in China from 1978-2022. *Journal of Global Health*. 2020;10(1):010803. doi:10.7189/jogh.10.010803
- [40] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x
- [41] World Health Organization. *Global Spending on Health 2023: Navigating the Fiscal Turbulence*. WHO; 2023. Available from: <https://www.who.int/publications/i/item/9789240074767>