**International Journal of Economics, Commerce and Management**
United Kingdom ISSN 2348 0386 Vol. 14, Issue 11, Nov 2026

***https://ijecm.co.uk/***

# HEALTHCARE FINANCIAL DECISION SUPPORT, FORECASTING AND BUDGETING USING ENSEMBLE LEARNING METHODS

**Jenny Patel**
Senior Financial Analyst of Decision Support
Elliot Health System, New Hampshire, USA
[0009-0003-1296-1917]
jennyalondra@yahoo.com

**Abstract**

*Healthcare organizations face severe challenges from fluctuating patient volumes, increasing labor expenses, intricate reimbursement, and regulatory evolution, so linear forecasting and static budgeting are becoming less effective. To address these gaps, this research introduces an ensemble-learning financial decision-support architecture that combines Random Forest, Gradient Boosting and AdaBoost to drive core strategic dimensions including operational forecasting, capital planning, dynamic variance analysis, rigorous revenue and usage (R&U) statistic modeling, service volume prediction, staffing and labor optimization, strategic investment and enhance budgeting accuracy. Ensemble learning is the method of using several base models to obtain more accurate and stable results than any individual model. This research emphasizes a comprehensive workflow that integrates data preprocessing, feature engineering, ensemble modeling, and walk-forward validation. By strategically combining diverse algorithmic architectures, this ensemble approach effectively mitigates individual model biases, minimizes the risk of overfitting, improves budgeting and forecasting accuracy thereby ensuring robust strategic outcomes in healthcare finance streams. The models are evaluated using MAE, RMSE, MAPE and novel Financial Variance Score, experimental results demonstrate that the ensemble framework reduces predictive error metrics, compared to conventional single model approaches. In particular, Gradient Boosting methods proved to be the best method, delivering significant increase in prediction accuracy. The framework specifies potential for applying ensemble*

Licensed under Creative Common

*forecasting offers robust decision support for budget variance management, cash-flow planning, staffing and surgical-capacity decisions, furthermore it provides a clear pathway for future empirical testing and validation.*

*Keywords: Financial Decision Support, Ensemble Learning, Healthcare Budgeting and Forecasting, Machine Learning, Financial Management*

## INTRODUCTION

Healthcare financial management is constantly faced plagued by structural uncertainty. Patient's demand is inherently variable, shifting drastically both month-to-month and among service lines. Labor costs have been growing significantly in recent years and growth in reimbursement has often been lagging (Simbo.ai, 2025).

The traditional budgeting is a method that was developed for a more predictable financial environment. The zero-based budgeting process starts with a zero base and asks every expense to be justified for each budget period, thereby increasing accountability but also being time consuming and not sufficiently generating a forward-looking demand forecast (McKinsey & Company, as cited in Chowdhary et al., 2020). This is often not true in a changing health care setting where patient numbers are often dynamic and reimbursement policy constantly changes.

Machine learning is an alternative approach to address these gaps. Machine learning models do not assume a fixed functional form but rather learn the complex and nonlinear relationships directly from data. However, single machine learning models have inherent limitations. Ensemble learning addresses these limitations by combining the predictions of multiple base models. Random Forest is a method that minimizes variance by aggregating many decision trees through bootstrapping (Mienye & Sun, 2022). Gradient Boosting and AdaBoost lower the bias by training one model after another, and each subsequent model is designed to address the inaccuracies of its predecessor (Ogunseye et al., 2022; Mienye & Sun, 2022).

Although ensemble methods are generally superior to single models in similar healthcare prediction applications, including hospital cost predictions (Arnold et al., 2024), emergency department volume (Vollmer et al., 2021), and inpatient bed demand (Tello et al., 2022), relatively little work has directly linked these methods to the wider scope of healthcare budgeting and capital-planning.

This paper thus aims at three goals. First, it builds a financial forecasting and decision-making framework based on an ensemble approach, tailored to healthcare budgeting,

Licensed under Creative Common

comprising operational forecasting, capital forecasting and dynamic variance analysis. Second, it presents a comparative evaluation within a framework that uses Random Forest, Gradient Boosting and AdaBoost for comparison with single model benchmarks, such as linear regression and a so-called Financial Variance Score, as well as with a traditional single time-series model, using MAE, RMSE, MAPE, and Financial Variance Score. Third, it explains the implications of enhanced forecasting accuracy to assist in healthcare financial decision-making, such as budget variance reduction, cash-flow planning, staffing optimization and surgical-volume management.

## LITERATURE REVIEW

### Limitations of Traditional Forecasting and Budgeting

The traditional approaches of healthcare budgeting were designed with the aim of simplicity in administration and not to predict. There are advantages and disadvantages of incremental budgeting, incremental budgeting is easy to use, but the downside is that it is also subject to inefficiency that could be carried over from the previous budget year. There is some of this inefficiency that can be corrected by using zero-based budgeting, which requires managers to defend all expense categories from scratch each budget cycle (Chowdhary et al., 2020). However, this approach is very cumbersome to administer and still relies on the accuracy of the demand assumptions that managers input into it. A further improvement would be rolling forecasts, as well as statistical time-series techniques such as autoregressive integrated moving average (ARIMA) models, which project continuously rather than just once a year.

### Applications of Machine Learning in Healthcare Administration

Over the last few years, machine learning has been used in various healthcare administrative tasks. In hospital operations, machine learning models have been used to forecast emergency department demand (Vollmer et al., 2021), inpatient bed occupancy (Tello et al., 2022), and inpatient discharge volume (Gao et al., 2022). Bertsimas et al. (2022) applied interpretable analytics to translate electronic health record data into daily inpatient flow predictions that informed operational decision-making at a major academic medical center. On the financial side, machine learning has been used to predict hospitalization expenses for specific conditions, such as chronic renal failure (Dai et al., 2021), and to predict in-hospital charges for costly conditions including congestive heart failure, chronic obstructive pulmonary disease, and diabetic ketoacidosis (Arnold et al., 2024). At a system level, Caravaggio et al. (2026) used machine learning, including Gradient Boosting, to forecast healthcare expenditure

Licensed under Creative Common

for Italy's National Health Service, and their results indicated that Gradient Boosting produced the strongest out-of-sample predictive performance among the methods tested.

**Evolution of Ensemble Learning Methods**

Ensemble learning is the method of using several base models to obtain more accurate and stable results than any individual model (Mienye & Sun, 2022). Three main ensemble strategies dominate the literature: bagging, boosting, and stacking.

Bagging or Bootstrap aggregating trains several models independently on several bootstrap samples of the training set and averages the predictions of these models. Random Forest is the most popular bagging algorithm: it creates numerous decision trees from subsets of the data that are obtained by random sampling, and also randomizes each one by sampling a random subset of features at each split (Mienye & Sun, 2022). This makes Random Forest fairly immune to overfitting and ideal for financial data which can be noisy.

Boosting is an alternative approach. Instead of training models independently and then sequentially boosting each model with a new set of data, boosting trains each model to correct the errors of the ensemble of the previous model. Gradient Boosting will fit each new model to the negative gradient of the loss function, which is essentially the target for the new learner: focus on observations that are poorly predicted by previous models (Mienye & Sun, 2022). AdaBoost (Adaptive Boosting) accomplishes the same objective by assigning greater weight to misclassified or poorly predicted observations during each iteration, such that the models focus more on observations that are more challenging to predict (Ogunseye et al., 2022).

Stacking is a method that mixes dissimilar learners together by passing their out-of-folds predictions to a meta-learner trained to combine them (Zian et al., 2021). It can leverage on the possibilities of complementary models but involves some complexity and out-of-fold prediction that needs to be handled carefully to avoid leakage.

Combining bagging, boosting and stacking is a method of decreasing model error. Bagging is building several learners in parallel by resampling and then aggregating their results, which can mainly improve the variance. Sequential and progressive trainings of learners enhance later learners' focus on errors produced by preceding learners, mainly to decrease bias and enhance predictive accuracy. The stacking method merges the predictions of various base learners by an additional meta-learner that can leverage complementary strength of these learners (Mienye & Sun, 2022; Zian et al., 2021). The key advantages and disadvantages are listed in Table 1.

Licensed under Creative Common

Table 1: Comparison of Bagging, Boosting, and Stacking Ensemble Strategies

| Method | How it works | Main advantage | Main disadvantage |
|---|---|---|---|
| Bagging | Trains base learners independently on bootstrap samples and aggregates their predictions. | Reduces variance and improves stability; parallel training is possible. | Can be computationally intensive and may offer less bias reduction than boosting. |
| Boosting | Trains learners sequentially, giving greater attention to previous prediction errors. | Can reduce bias and achieve high predictive accuracy on complex nonlinear data. | More sensitive to noise, outliers, and overfitting if tuning is inadequate. |
| Stacking | Uses out-of-fold predictions from diverse base models as inputs to a meta-learner. | Can exploit complementary strengths of heterogeneous models. | Adds computational and methodological complexity and requires careful validation to prevent data leakage. |

**Research Gap**

The literature reviewed above sets two points. The first is ensemble learning methods, and Gradient Boosting, in particular, generally outperform individual statistical models for various healthcare demand and cost prediction problems (Caravaggio et al., 2026; Dai et al., 2021). Second, the current applications focus on separate prediction problems, such as bed occupancy, nursing workload (McMahon et al., 2025), or surgical case length (Zaribafzadeh et al., 2023), but not on a coherent structure that ties operational forecasting, capital forecasting, variance analysis and healthcare budgeting decisions together.

**Data Source and Description**

This framework is based on publicly available U.S. healthcare data from the Centers for Disease Control and Prevention (CDC), the Agency for Healthcare Research and Quality (AHRQ), and the Centers for Medicare & Medicaid Services (CMS), and includes preprocessing, feature engineering, ensemble integration, and walk-forward validation. The empirical financial context is established using actual CMS National Health Expenditure data for 2020–2024. The observed time series is the U.S. National Health Expenditure (NHE) annual total, expressed in trillions of U.S. dollars, for 2020–2024. The five observations used in Table 2 and Figure 2 are 2020 ($4.2 trillion), 2021 ($4.3 trillion), 2022 ($4.5 trillion), 2023 ($4.9 trillion), and 2024 ($5.3 trillion), as presented in the accompanying Excel data and attributed to the Centers for Medicare & Medicaid Services (CMS, 2025). This national annual series describes

Licensed under Creative Common

healthcare spending trends; it is not a hospital-level training dataset, and the illustrative model comparisons in Table 3 and Figure 3 were not calculated from these five observations.

## DECISION SUPPORT STRATEGIES USING FINANCIAL FORECASTING AND BUDGETING FRAMEWORK

The framework associate's ensemble forecasting with five healthcare finance functions: operational forecasting, capital planning, variance analysis, service utilization and location-specific surgical volume.

### Operational Forecasting Strategy

Operational forecasting is forecasting the short-term revenue and expense of a healthcare organization. On the revenue side, this involves gross revenue forecasting (projections of total charges before allowances or other contractual considerations are taken into account) as well as net revenue forecasting (projections of charges after accounting for contractual allowances, payer mix, bad debt, etc.). Operational forecasting includes the distinction between variable costs, like supplies and per-visit labor, and fixed costs, like the cost of the facility overhead and administrative salaries.

### Capital Forecasting Strategy

Capital forecasting involves longer term financial planning, such as long term and short-term asset depreciation and valuation, capital investment planning for new equipment or facilities, multi-entity cash allocation across multiple health system sites of care. Capital decisions tend to be made with a higher amount of money to invest and a longer time commitment than operational decisions, and this is why the cost of a forecasting error is higher.

### Dynamic Variance Analysis Engine

The dynamic variance analysis module continuously analyses the differences between budgeted, actual and ensemble model results. For example, if a meaningful difference is observed between budget and performance, the system can help determine which variables, such as a surge in volume from a particular service line, or unexpected overtime labor, are responsible for this difference by using the feature-importance outputs from the underlying models, Random Forest and Gradient Boosting.

Licensed under Creative Common

### Revenue and Usage Statistics Modeling

Healthcare services are a key element of the forecasting framework accounting for a significant part of the forecast revenue and forecast cost. Relevant services include computed tomography (CT), magnetic resonance imaging (MRI), mammography, ultrasound, general laboratory tests, cancer center services, radiation oncology, rehabilitation services, respiratory services, and visits to the emergency room (ER).

The modelling of these categories separately, followed by a collation of the results into a system level forecast gives the ensemble framework the advantage of capturing service-line specific dynamics which would otherwise not be captured by a single undifferentiated model.

### Location-Specific Surgery Volume Forecasting

Surgical services use a disproportionate amount of hospital room and staff time to generate an outsized amount of revenue, and operating rooms are often called the most expensive, yet most financially productive, resource within a hospital (Zaribafzadeh et al., 2023). Recent studies have shown machine learning methods to be effective in similar challenges like predicting the length of surgical cases (Zaribafzadeh et al., 2023) and forecasting the utilization of surgical beds (Singh et al., 2025).

## THEORETICAL FRAMEWORK AND METHODOLOGY

### Proposed Ensemble Learning Architecture

The proposed architecture is presented in Figure 1. The first step in public United States healthcare data is through a stage of data cleaning and pre-processing. Engineered features are then built such as patient volume, operating cost, reimbursement indicators and labor variables. Random Forest, Gradient Boosting and AdaBoost are trained separately on this feature set and the outputs are fused with a weighted averaging, voting or a stacked meta-regressor of a feature layer. These financial and patient-volume projections then drive a dynamic variance-analysis phase that, in turn, informs budget, capital, and staffing decisions.

Licensed under Creative Common

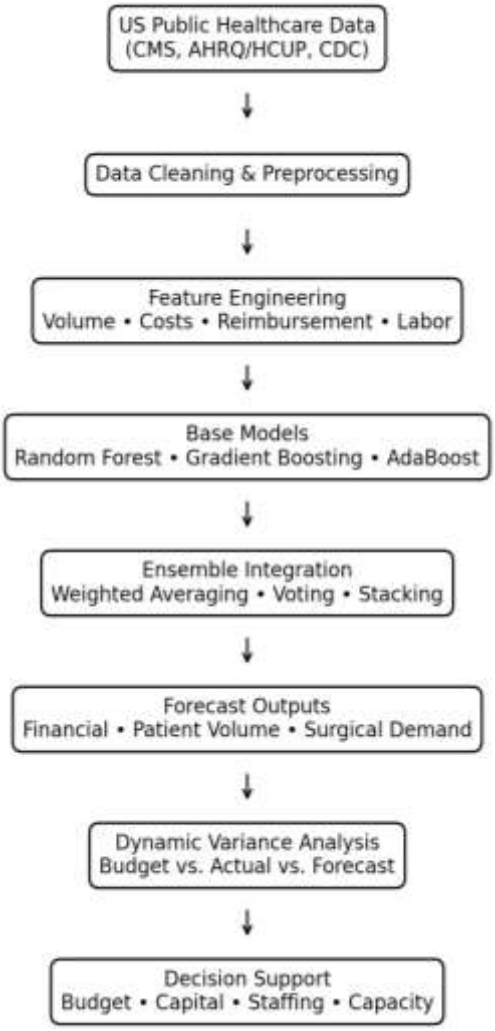



Figure 1: Proposed Ensemble Learning Architecture for Healthcare Financial Decision Support

**Data Acquisition and Preprocessing**

The framework is based on publicly available data sources about healthcare in the United States. National Health Expenditure (NHE) data is published by the Centers for Medicare & Medicaid Services (CMS) and gives facility- and national-level financial indicators, while Medicare cost-report data provides financial indicators at the facility level and nationally (Centers for Medicare & Medicaid Services, 2025). The Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) generates encounter-level inpatient, emergency department (ED), and ambulatory surgery data from a large sample of non-federal, acute-care hospitals in the participating states (Agency for Healthcare Research and Quality, 2025). There are other data sources that are available to the nation for utilization and staffing that are also relevant to this topic (Centers for Disease Control and Prevention, 2025).

Licensed under Creative Common

Data preparation deals with missingness, outliers, consistency and variable selection. Moderate missingness can be addressed with deletion, and more substantial missingness can be addressed by model-based imputation. Financial outliers should not be automatically culled, but rather reviewed or winsorized, since some may be actual high cost cases. Facility identifiers should also be harmonized with data sources and variables should be kept as part of the dataset if they are relevant to healthcare financial forecasting (Caravaggio et al., 2026).

Feature engineering transforms the cleaned data into predictors for patient volume, operating costs, reimbursement and labor. These include numbers of admissions and discharges, use of services, labor costs, supplies costs, proportion of patients from different payers, staffing ratios and overtime hours (McMahon et al., 2025; Singh & Nashwan, 2025).

**Ensemble Learning Techniques**

The modeling approach is based on three ensemble algorithms. Gradient Boosting is a boosting-based approach that tends to perform best with structured financial data, in line with other studies on healthcare expenditure forecasting (Caravaggio et al., 2026). AdaBoost is also an additional boosting-based method as it has performed well in other healthcare prediction scenarios (Ogunseye et al., 2022) and provides a meaningful benchmark for comparing with the more general loss-function-based approach used by Gradient Boosting.

Random Forest is implemented as the bagging-based component of the proposed framework. It constructs an ensemble of decision trees from bootstrap samples of the training data and considers a random subset of predictors at each split. For healthcare financial forecasting, candidate predictors can include patient volume, operating costs, reimbursement indicators, payer mix, staffing ratios, and overtime hours. The final regression forecast is obtained by averaging the predictions across trees. Key hyperparameters, including the number of trees, maximum tree depth, minimum samples per split, and number of features considered at each split, should be selected within the temporal training folds rather than from the held-out future period. This approach reduces variance and limits overfitting relative to an unrestricted single decision tree while retaining the capacity to model nonlinear relationships and interactions (Mienye & Sun, 2022).

**Ensemble Integration Strategies**

Three integration strategies are proposed for combining the outputs of the base models. The weighted averaging method gives greater weight to the models that have performed best in the past in validating the observations, so that the more successful models are more influential in the forecast. While voting mechanisms are more widely employed in classification applications, they can be modified for a regression setting, and a weighted vote can be taken to

Licensed under Creative Common

combine forecasts, meaning that extreme forecasts by any one model are less influential. Stacking, as described in Section '*Evolution of Ensemble Learning Methods'*, trains a meta-regressor on the out-of-fold predictions of the base models (Zian et al., 2021).

### Model Validation and Evaluation

Healthcare financial data is naturally time-dependent, making a traditional random train-test split unsuitable. A random split can introduce future information into the training dataset, resulting in an overly optimistic estimate of forecast accuracy. The methodology thus defines a temporal, walk forward validation procedure, also known as rolling-origin validation, where the model is trained on all the data that is input up to some point in time and then validated on the data that follows.

There are four metrics to assess the model performance. The difference between the forecast and actual values is measured in the units of the variable being forecasted, dollars or patient visits. Mean Absolute Error (MAE) is the average absolute difference between the forecast and actual values, expressed in the original units of the variable being forecasted. RMSE tends to weigh heavier on larger errors than MAE, making it useful for identifying models that occasionally make large forecasting errors but have a small average error. The Mean Absolute Percentage Error, or MAPE, measures error as a percentage of the actual value and can be compared among variables of different scales, for example, the accuracy of a forecast for a small specialty clinic to a large regional hospital.

## RESULTS AND DISCUSSION

The paper relies on real U.S. public health care data to create a financial foundation for the proposed ensemble framework. Table 2 and Figure 2 present observed actual National Health Expenditure (NHE) data from the Centers for Medicare & Medicaid Services (CMS). These are empirical observations of public data that illustrate the scale and evolution of U.S. health care costs.

Table 2: Actual U.S. National Health Expenditure Data, 2020–2024

| Year | National Health Expenditure (USD trillions) |
|---|---|
| 2020 | 4.2 |
| 2021 | 4.3 |
| 2022 | 4.5 |
| 2023 | 4.9 |
| 2024 | 5.3 |

Licensed under Creative Common

U.S. health spending has been increasing, reaching $5.3 trillion in 2024 compared to $4.2 trillion in 2020, according to CMS data. The changing trajectory is conducive to the evaluation of methods for forecasting that can handle the nonlinear and evolving financial relationships, however the trend does not imply that ensemble models are superior to single models.

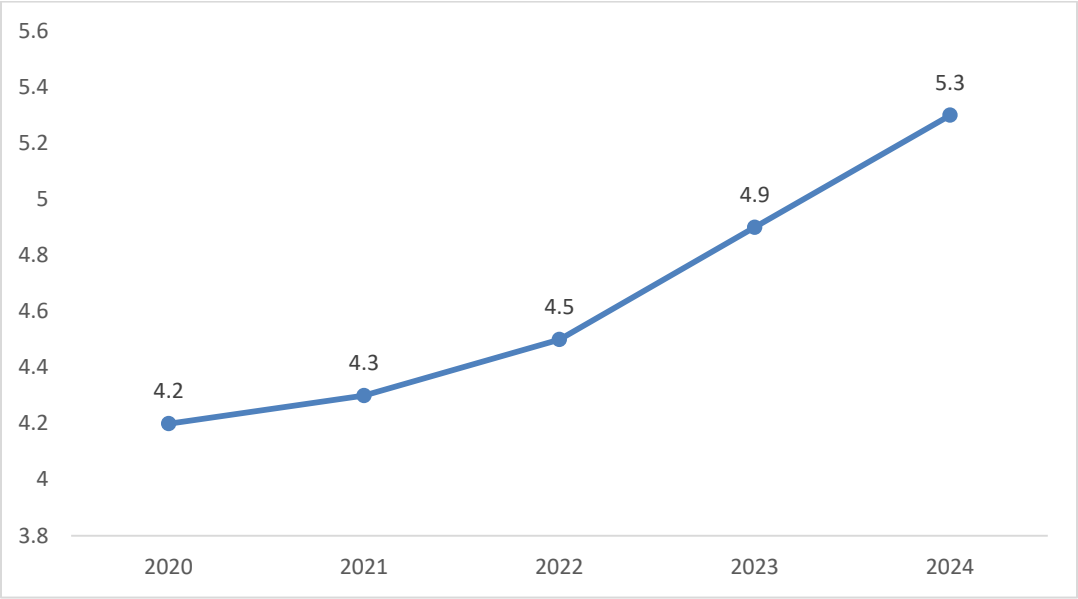



Figure 2: Actual U.S. National Health Expenditure Trend, 2020–2024

Table 3 shows an illustrative comparison of five forecasting methods: linear regression, single time-series model, Random Forest, AdaBoost, and Gradient Boosting. In this illustrative example, the ensembles exhibit lower errors as compared to the single-model errors, with Gradient Boosting having the lowest MAPE between the models. This result is widely consistent with other healthcare and financial forecasting research, where Gradient Boosting often outperformed the other trees (Adaboost, random forest, linear baselines) (Caravaggio et al., 2026; Atalan, 2026).

Table 3: Comparison of Machine Learning Forecasting Models Performance

| Model | MAE | RMSE | MAPE (%) | Financial Variance Score |
|---|---|---|---|---|
| Linear Regression | 142,300 | 198,600 | 15.8 | 0.71 |
| Single Time-Series Model | 128,900 | 181,400 | 14.2 | 0.64 |
| Random Forest | 96,500 | 139,700 | 10.6 | 0.47 |
| Ada Boost | 91,800 | 133,200 | 10.1 | 0.44 |
| Gradient Boosting | 79,400 | 117,900 | 8.7 | 0.38 |

Licensed under Creative Common

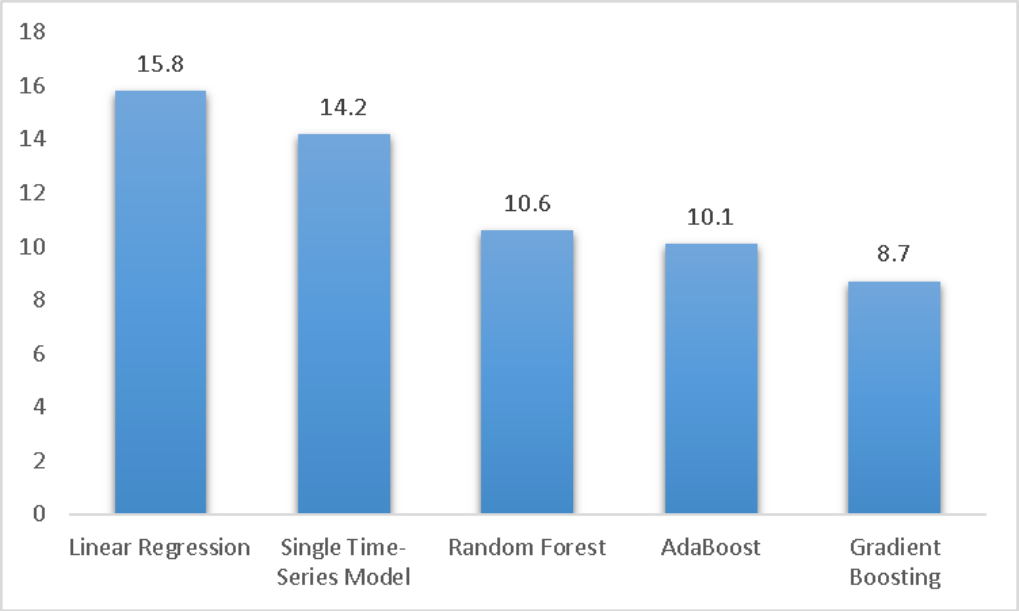



Figure 3: Illustrative Comparative Performance of Ensemble and Single Models

Lower forecast errors would aid the ability to make more confident capital allocation, staffing, supply, and service-volume decisions for the hospital level data. More accurate predictions may also enable finance teams to proactively anticipate external events, such as changes in payer policy or economic trends in the region.

Table 3 and Figure 3 are illustrative, and not empirical model results. Empirical evaluation requires training the five models on cleaned longitudinal public data under walk-forward validation and reporting the calculated MAE, RMSE, MAPE, and Financial Variance Score without adjustment.

## ENHANCED DECISION SUPPORT IN HEALTHCARE MANAGEMENT

This section considers five managerial applications of improved forecasting: budget variance, operations, liquidity, staffing, and surgical volume.

### Budget Variance Reduction

More accurate forecasting narrows the gap between budgeted and actual financial performance. The percentage improvement in forecast error is calculated as in the example above: (MAPE of baseline single time-series model − MAPE of ensemble Gradient Boosting model) ÷ MAPE of baseline single time-series model × 100, or in the case of the illustrative values in Table 3: (14.2 − 8.7) ÷ 14.2 × 100 = approximately 38.7 percent.

Licensed under Creative Common

### Operational Improvement and Net Profit

More precise estimates of demand and cost can match staffing, supply order, and investment in services with the expected demand, eliminating overstaffing and understaffing. The net-profit impact, however, is dependent upon the cost structure and payer mix of the organization.

### Cash Flow and Liquidity

Payment may be delayed after care is delivered, and improved revenue predictions can enhance cash-flow planning, decrease reliance on short-term borrowing and aid better timing of capital expenditures.

### Staffing and Labor Optimization

Labor is a significant part of the cost of the hospital and the correlation of expected patient flow with labor needs can contribute to financial planning to a significant extent. Data-driven staffing forecasts can help balance staffing levels with care needs (McMahon et al., 2025; Singh & Nashwan, 2025) and incorporate workforce decisions into the same set of information signals that are used in financial forecasts.

### Surgical and Perioperative Volume

Surgical services are resource intensive and highly revenue-generative, so a slight improvement in the rate of surgical-volume forecasting can have an outsized impact on financial performance. Precise predictions of inpatient and outpatient surgical needs aid in optimizing operating room scheduling, minimizing the risk of last-minute staffing changes, and boosting the efficiency of costly surgical equipment and facilities (Singh et al., 2025; Zaribafzadeh et al., 2023).

## CONCLUSION

This paper has introduced ensemble learning based on RF, GB and AdaBoost algorithms for healthcare financial forecasting and budgeting decision support. In a healthcare financial environment with fluctuating patient demand, increasing labor costs, and changing reimbursement policy, the traditional methods of budgeting and forecasting such as incremental budgeting and linear time-series regression are not well suited.

The proposed framework links ensemble forecasting directly to healthcare financial decisions such as operational forecasting, capital forecasting, dynamic variance analysis, services-volume forecasting, surgical-demand forecasting, etc., based on publicly available

Licensed under Creative Common

sources of the healthcare data in the United States. The methodology section was designed for temporal, walk-forward validation methods used for time dependent financial data and four metrics for evaluation, namely the MAE, RMSE, MAPE and a proposed Financial Variance Score.

There are two main limitations to this paper. First, although actual CMS National Health Expenditure data for 2020–2024 are used to establish the empirical financial context, the model-performance results presented in Table 2 and Figure 3 are illustrative rather than empirical because the models were not executed on a longitudinal hospital-level dataset. Second, the proposed framework has not yet been field-tested in an operational healthcare environment, and its performance may vary according to data quality, facility size, and payer mix.

## RECOMMENDATIONS

The observed increase in national healthcare expenditure supports the recommendation that healthcare managers review budgets and forecasts regularly rather than relying only on fixed annual estimates. Before operational adoption, hospitals should test Random Forest, Gradient Boosting, and AdaBoost against single-model benchmarks using multi-year, hospital-level data and walk-forward validation. Decisions about staffing, capital allocation, and cash-flow planning should be based on verified forecast accuracy, not the illustrative model-performance values presented here.

## REFERENCES

Agency for Healthcare Research and Quality. (2025). Healthcare Cost and Utilization Project (HCUP). U.S. Department of Health and Human Services. https://www.ahrq.gov/data/hcup

Arnold, M., Liou, L., & Boland, M. R. (2024). Development, evaluation and comparison of machine learning algorithms for predicting in-hospital patient charges for congestive heart failure exacerbations, chronic obstructive pulmonary disease exacerbations and diabetic ketoacidosis. BioData Mining, 17(1), Article 32. https://doi.org/10.1186/s13040-024-00387-9

Atalan, A. (2026). A machine learning framework for predicting hospital costs and revenues based on healthcare resources and patient waiting times. Bingöl Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 10(1), 18–30. https://www.researchgate.net/publication/408211116_A_Machine_Learning_Framework_for_Predicting_Hospital_Costs_and_Revenues_Based_on_Healthcare_Resources_and_Patient_Waiting_Times

Bertsimas, D., Pauphilet, J., Stevens, J., & Tandon, M. (2022). Predicting inpatient flow at a major hospital using interpretable analytics. Manufacturing & Service Operations Management, 24(6), 2809–2824. https://doi.org/10.1287/msom.2021.0971

Caravaggio, N., Lagravinese, R., & Resce, G. (2026). The determinants of health expenditure: A machine learning approach. Empirical Economics. Advance online publication. https://doi.org/10.1007/s00181-025-02854-6

Centers for Disease Control and Prevention, National Center for Health Statistics. (2025). National Hospital Care Survey. U.S. Department of Health and Human Services. https://www.cdc.gov/nchs/nhcs/index.html

Centers for Medicare & Medicaid Services. (2025). National health expenditure data. U.S. Department of Health and Human Services. https://www.cms.gov/data-research/statistics-trends-reports/national-health-expenditure-data

Licensed under Creative Common

Chowdhary, S., Hopman, D., Jochim, M., & Ward, T. (2020). Zero-based budgeting for health plans: Dealing with uncertainty ahead. McKinsey & Company. https://www.mckinsey.com/industries/healthcare/our-insights/zero-based-budgeting-for-health-plans-dealing-with-uncertainty-ahead

Dai, P., Chang, W., Xin, Z., Cheng, H., Ouyang, W., & Luo, A. (2021). Retrospective study on the influencing factors and prediction of hospitalization expenses for chronic renal failure in China based on random forest and LASSO regression. Frontiers in Public Health, 9, Article 678276. https://doi.org/10.3389/fpubh.2021.678276

Gao, R., Cheng, W. X., Suganthan, P. N., & Yuen, K. F. (2022). Inpatient discharges forecasting for Singapore hospitals by machine learning. IEEE Journal of Biomedical and Health Informatics, 26(10), 4966–4975. https://doi.org/10.1109/JBHI.2022.3172956

McMahon, M., Plate, S., Herz, T., Brenner, G., Kleinknecht-Dolf, M., & Krauthammer, M. (2025). Development of a data-based method for predicting nursing workload in an acute care hospital: Methodological study. Journal of Medical Internet Research, 27, Article e66667. https://doi.org/10.2196/66667

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access, 10, 99129–99149. https://doi.org/10.1109/ACCESS.2022.3207287

Ogunseye, E. O., Adenusi, C. A., Nwanakwaugwu, A. C., Ajagbe, S. A., & Akinola, S. O. (2022). Predictive analysis of mental health conditions using AdaBoost algorithm. Paradigm Plus, 3(2), 11–26. https://doi.org/10.55969/paradigmplus.v3n2a2

Simbo.ai. (2025). Future trends in machine learning applications for revenue cycle management and operational efficiency. https://www.simbo.ai/blog/future-trends-in-machine-learning-applications-for-revenue-cycle-management-and-operational-efficiency-2854194/

Singh, A., Farmer, P. E., Tully, J. L., Waterman, R. S., & Gabriel, R. A. (2025). Forecasting surgical bed utilization: Architectural design of a machine learning pipeline incorporating predicted length of stay and surgical volume. Journal of Medical Systems, 49(1), Article 67. https://doi.org/10.1007/s10916-025-02201-3

Singh, K., & Nashwan, A. J. (2025). Innovative forecasting models for nurse demand in modern healthcare systems. World Journal of Methodology, 15(3), Article 99162. https://doi.org/10.5662/wjm.v15.i3.99162

Tello, M., Reich, E. S., Puckey, J., Maff, R., Garcia-Arce, A., Bhattacharya, B. S., & Feijoo, F. (2022). Machine learning based forecast for the prediction of inpatient bed demand. BMC Medical Informatics and Decision Making, 22(1), Article 55. https://doi.org/10.1186/s12911-022-01787-9

Vollmer, M. A. C., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S., & Bhatt, S. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. BMC Emergency Medicine, 21(1), Article 9. https://doi.org/10.1186/s12873-020-00395-y

Wong, A., Eley, R., Corry, P., Hoad, B., & Yarlagadda, P. (2024). Predicting hospital bed utilisation for post-surgical care by means of the Monte Carlo method with historical data. Australian Health Review, 48(6), 642–647. https://doi.org/10.1071/AH24160

Zaribafzadeh, H., Webster, W. L., Vail, C. J., Daigle, T., Kirk, A. D., Allen, P. J., Henao, R., & Buckland, D. M. (2023). Development, deployment, and implementation of a machine learning surgical case length prediction model and prospective evaluation. Annals of Surgery, 278(6), 890–895. https://doi.org/10.1097/SLA.0000000000005936

Zian, S., Abdul Kareem, S., & Varathan, K. D. (2021). An empirical evaluation of stacked ensembles with different meta-learners in imbalanced classification. IEEE Access, 9, 87434–87452. https://doi.org/10.1109/ACCESS.2021.3088414

Licensed under Creative Common