

HealthFlow AI: A Validated Dual-Target Forecasting System with Capacity-Guarded Resource Estimation and Hospital-Scoped Access Control for Real-World Hospital Operations

Mrs. Shivaranjini¹, Suparn Nayak², Varsha K N³, Parth Pandith⁴, Lingaraj V Patil⁵

¹Assistant Professor, Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India

^{2,3,4,5}Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India

Abstract - Hospital operations forecasting systems are frequently evaluated on synthetic or idealized data, leaving open whether reported gains survive contact with the noise, missingness, and reporting-policy shifts present in real administrative healthcare data. This paper presents HealthFlow AI, a dual-target hospital forecasting and resource-planning platform, and reports its validation on a real, publicly available dataset: 309,060 weekly hospital-level records spanning 1,489 U.S. hospitals from December 2019 to April 2024, drawn from the U.S. Department of Health and Human Services' hospital capacity reporting program. The system forecasts two independently modeled targets one to four weeks ahead patient admissions (demand) and inpatient bed occupancy (resource utilization) using a residual-correction gradient-boosted architecture layered on a four-week moving-average baseline, together with quantile models that yield calibrated 80% prediction intervals (observed coverage 78.9-82.9% across all horizons). Contrary to a naive expectation of uniform improvement, we report an honest negative result: the learned admissions model does not outperform the simple moving-average baseline at any horizon, while the occupancy model outperforms it at every horizon (mean absolute error 6.62-7.60 beds/week versus 6.78-7.86 for the baseline). We further describe a capacity-guard methodology that detects and corrects two distinct real-world data-quality failure modes in hospital-capacity reporting zero/missing capacity values and anomalously low-but-positive capacity values the latter of which we trace to a documented U.S. federal policy change (the May 2023 expiration of the COVID-19 Public Health Emergency) that altered what facilities were required to report. Finally, we describe a hospital-scoped role-based access control layer with audit logging and refresh-token rotation, and a large-language-model explanation agent hardened against generating fabricated analyses when forecast data is unavailable. All reported results are drawn from held-out, chronologically split test data and cross-verified across independent re-runs of the training and evaluation pipeline.

Index Terms- Hospital admission forecasting, gradient boosting, LightGBM, resource capacity planning, healthcare data quality, role-based access control,

large language model agents, explainable artificial intelligence, healthcare operations.

I. INTRODUCTION

Short-term forecasting of hospital admissions and bed occupancy supports staffing, patient-flow, and capacity decisions. Recent studies address weekly and daily occupancy, admissions forecasting, epidemic-driven demand, and operational deployment [1]-[7], [11]-[15]. Yet strong retrospective accuracy alone does not guarantee reliable use: administrative data can contain missing or suppressed values, reporting rules can change, and performance can drift as populations and practice evolve. A useful system must therefore be evaluated against simple baselines on chronological holdouts and must expose data-quality and uncertainty limitations.

This paper reports the redesign and empirical validation of HealthFlow AI, a hospital operations intelligence platform, against a real public dataset: the U.S. Department of Health and Human Services' (HHS) weekly hospital capacity and utilization reporting data. We deliberately report results that did not turn out as initially hoped, because we consider the negative result and the process of distinguishing a genuine model limitation from a data-quality artifact to be as valuable a contribution as the cases where the model succeeds.

The contributions of this work are as follows:

1) We validate a horizon-specific, residual-correction gradient-boosted forecasting architecture on 309,060 real weekly hospital-level records across 1,489 hospitals, forecasting two independent targets (admissions and inpatient bed occupancy) one to four weeks ahead, with calibrated quantile-based prediction intervals.

2) We report an honest comparative finding: the learned model does not outperform a simple four-week moving-average baseline for admissions forecasting at any evaluated horizon, while it does outperform the same baseline for occupancy forecasting at every horizon and we analyze why the two targets behave differently.

3) We describe a capacity-guard data-quality methodology that identifies and corrects invalid (zero or missing) capacity reports, and separately flags anomalously low but technically valid capacity reports; we trace the latter, in a specific case, to a documented change in federal reporting requirements following the expiration of the COVID-19 Public Health Emergency declaration.

4) We describe a hospital-scoped role-based access control (RBAC) architecture with per-hospital data isolation, JWT-embedded access claims, refresh-token rotation, and audit logging verified through automated integration tests including adversarial mixed-authorization requests.

5) We describe a safeguard added to the platform's large-language-model (LLM) explanation agent that prevents it from generating fabricated clinical-sounding analysis when the underlying forecast data does not exist, a failure mode we discovered during our own testing.

II. RELATED WORK

The 2026 literature emphasizes operational relevance and temporal robustness. Porter et al. [1] prospectively forecast pediatric transplant and cellular-therapy census 15 to 30 days ahead by combining planned admissions, predicted unplanned admissions, and discharge probabilities. Williams et al. [2] examine covariate and concept drift in emergency-triage admission prediction and report that periodic retraining improves performance over long test periods. Palacios et al. [3] build short-term bed-saturation risk forecasts from ensembles at two hospitals, linking forecast uncertainty to decisions. Fattouh et al. [4] evaluate Prophet on daily occupancy records over horizons from 30 to 180 days, showing that interpretable time-series models can be competitive with more complex alternatives.

In 2025, Avinash and Mishra [5] combine Bayesian model averaging with deep sequence models for weekly mental-health bed occupancy, explicitly representing model uncertainty. Tobin et al. [6] describe a modular, incidence-informed approach to ward and ICU occupancy forecasts across Australian jurisdictions; forecast errors vary with epidemic phase and jurisdiction size. Mellor et al. [7] report operational two-week forecasts of COVID-19, influenza, and RSV admissions and bed occupancy in England, illustrating the value of calibrated ensembles and multiscale delivery. Zhang et al. [8] investigate gradient boosting together with GPT-2 for hospital-admission prediction. Lovett et al. [9] evaluate a pragmatic occupancy model whose error rises sharply during a changing pandemic period, underscoring the need for continuous monitoring. Avinash et al. [10] compare six machine-learning approaches for weekly occupancy in an Indian mental-health hospital and use forecast comparisons to inform capacity planning.

Three 2024 studies provide complementary comparisons. Porto and Fogliatto [11] evaluate feature-engineered machine-learning methods for emergency-department arrivals across multiple departments, supporting the use of lag and calendar information. Albrecht et al. [12] forecast severe respiratory-disease hospitalizations using machine-learning models, a setting in which changing disease mix and public-health conditions matter. Tuominen et al. [13] compare advanced machine-learning approaches for emergency-department occupancy and provide a direct operational analogue to the admissions and occupancy targets studied here.

Rostami-Tabar et al. [14] develop probabilistic machine-learning forecasts for daily COVID-19 admissions, emphasizing distributions rather than point predictions. Baas et al. [15] present real-time COVID-19 ward and ICU bed-occupancy forecasting, connecting epidemic dynamics with hospital capacity. Together, these studies motivate evaluating both point accuracy and interval coverage, testing against simple baselines, and treating reporting shifts as part of the deployment problem rather than as noise to hide.

HealthFlow AI differs in scope: it evaluates two weekly targets over 1,489 hospitals with a per-hospital chronological holdout, reports the learned model when it loses to a moving-average baseline, and adds explicit capacity-quality flags and hospital-scoped access controls. These results are an implementation and retrospective validation study, not evidence of prospective clinical effectiveness.

III. SYSTEM ARCHITECTURE

HealthFlow AI is organized as a serverless pipeline with four layers. A scheduling layer executes weekly retraining and forecast-generation jobs, replacing an earlier daily-cadence design once the underlying data was confirmed to be reported at weekly granularity; the previous daily job is retained as a documented backward-compatible delegator rather than removed outright. A PostgreSQL database (Neon) remains the single source of truth, with the forecasts table extended in this revision to store, per hospital and horizon: the target name (admissions or inpatient_beds_used), the point estimate, the 10th- and 90th-percentile quantile bounds, the derived resource gap, and a capacity_source field (reported, historical_median_fallback, or reported_anomalous_low) described in Section VI. A FastAPI serving layer exposes forecast, agent, and administrative endpoints, now uniformly enforcing hospital-scoped access control (Section VII) on every data-reading route. A Next.js frontend renders two role-differentiated views: a hospital-staff dashboard scoped to a user's assigned facility or facilities, and an administrative console with network-wide

analytics, hospital-management tooling, and an access-audit trail.

IV. DATASET

The evaluation dataset was constructed from HHS's public weekly hospital capacity and utilization reporting program by filtering the full national dataset to hospitals meeting two quality criteria: an average reporting coverage of at least six of seven days per week across their full history, and a minimum of 100 weeks of continuous reporting (approximately two years), ensuring sufficient history for lag and rolling-window feature construction. This produced 309,060 hospital-week records across 1,489 hospitals, spanning December 29, 2019 to April 21, 2024. Admissions values (approximated via reported emergency department visit counts) are non-null for 88.97% of records; inpatient bed occupancy values are non-null for 93.37% of records. The source data encodes suppressed small counts (a CMS privacy-protection convention for counts of one to three) with a sentinel value, which we convert to missing prior to feature construction; LightGBM's native handling of missing values is used rather than imputation for these cases. Hospitals were partitioned into three roughly equal size tiers (small: $n=496$, medium: $n=496$, large: $n=497$) by median admissions volume, for the tier-stratified evaluation reported in Section VIII.

V. METHODOLOGY

A. Forecasting Architecture

For each of the two targets (admissions, inpatient bed occupancy) and each forecast horizon h in $\{1, 2, 3, 4\}$ weeks, we compute a four-week trailing moving average (using only data available at or before the feature date, with no leakage) as a baseline forecast, and train a LightGBM regressor to predict the residual between the actual future value and this baseline. The final point forecast is the sum of the baseline and the predicted residual. This residual-correction design was adopted after an initial direct-regression approach (predicting the raw target value from lag and calendar features) failed to outperform even a naive persistence baseline; residual correction on top of the moving average recovered a meaningful improvement for one of the two targets, discussed in Section VIII-A. Feature inputs include lagged target values at 1, 2, 4, and 8 weeks; 4- and 8-week trailing rolling mean and standard deviation; hospital bed and ICU capacity; geographic coordinates derived from each facility's reported address; and categorical hospital, state, and facility-subtype identifiers, passed to LightGBM as native high-cardinality categorical

B. Quantile Models and Evaluation Protocol

In addition to the point-estimate residual model, we train quantile regressors at the 10th and 90th percentiles for each target and horizon, using the same residual-correction construction, yielding an 80% nominal prediction interval. Evaluation uses a strictly chronological, per-hospital train/test split: for each hospital, the most recent 15% of weeks are held out as test data, with all earlier weeks used for training; no random shuffling is used at any stage, and this was independently verified by confirming that the maximum training-set date precedes the minimum test-set date for every one of the 1,489 hospitals. Reported metrics are mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE), computed separately by hospital-size tier as well as overall, since a single blended metric across a roughly thirteen-fold range in hospital scale obscures materially different accuracy characteristics for small and large facilities.

C. Resource Gap and the Capacity Guard

A resource gap is computed as the forecasted inpatient bed occupancy minus 85% of the hospital's reported bed capacity, flagging weeks where a facility is projected to approach or exceed a conventional safe-operating threshold. During evaluation, this calculation initially produced an implausible maximum gap of 927.96 beds for a single facility, traced to that facility reporting a capacity of exactly zero for a period a known artifact of the underlying CMS reporting program's handling of missing submissions. We implemented a capacity guard: any hospital-week with reported capacity less than or equal to zero, or missing, is assigned that hospital's own historical non-zero median capacity instead, with the substitution recorded via a capacity_source flag rather than performed silently. Applying this guard reduced the affected facility's gap to a plausible range and reduced the proportion of test-set hospital-weeks flagged as exceeding the 85% threshold from 14.57% to 11.69% (Table III), with the mean gap among flagged weeks falling from 52.72 to 20.65 beds. A second, more subtle anomaly remained after this fix: a facility (UC San Diego Health Hillcrest) reporting a technically positive but implausibly low capacity of 73-119 beds against a multi-year historical median of 839.3 beds, sustained across 33 consecutive weeks beginning precisely the week following the May 11, 2023 expiration of the U.S. federal COVID-19 Public Health Emergency declaration, while its reported inpatient census remained at 682-765 beds throughout the same period. We interpret this as a reporting-scope change the facility's submitted capacity figure appears to have switched from whole-campus to a sub-unit count following the change in federal reporting requirements rather than a genuine 91% reduction in physical bed capacity with an unchanged patient census, which would be operationally implausible. We extended

the capacity guard with a second, non-corrective rule: any positive reported capacity below 35% of that hospital's own historical median is flagged with `capacity_source = reported_anomalous_low`, a threshold chosen empirically after observing that only 0.34% of all positive-capacity hospital-weeks nationally fall below this ratio, against 0.93% below a 50% threshold, indicating a genuine discontinuity in the data around this point rather than an arbitrary cutoff. Unlike the zero/missing case, this second rule flags rather than imputes the affected records, since the true capacity for these weeks cannot be reliably reconstructed; 289 of 41,992 test-set hospital-weeks (0.69%) were flagged under this rule, and the residual maximum reported gap of 660.57 beds is attributable to this flagged case rather than to an uncorrected defect in the pipeline.

Figure 1 summarizes the end-to-end forecasting procedure, including chronological evaluation and the explicit capacity-quality guard.

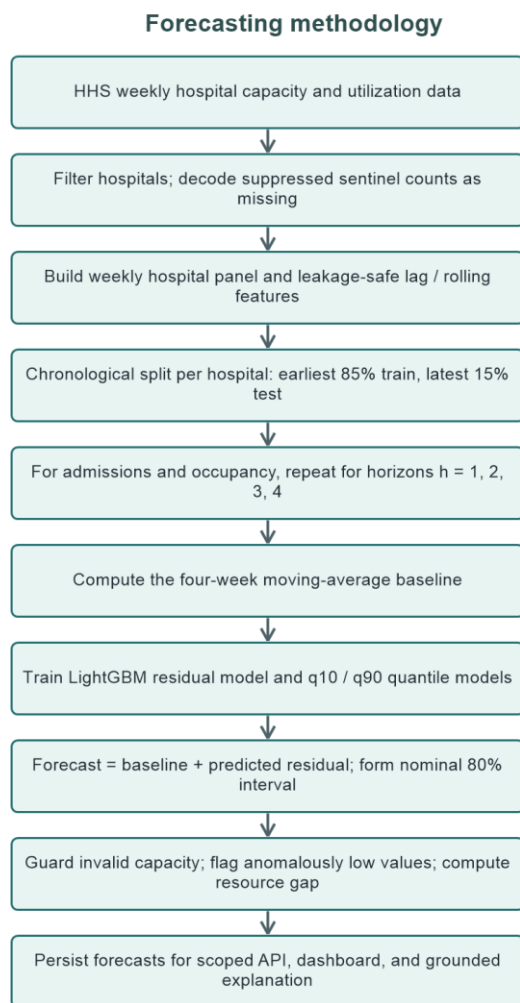


Fig. 1. HealthFlow AI forecasting methodology and capacity-guard workflow.

VI. HOSPITAL-SCOPED ACCESS CONTROL

Prior to this revision, any authenticated user could query forecast data for any hospital in the system, regardless of their organizational affiliation. We implemented role-based access control with two roles, admin and hospital_staff, backed by a join table mapping staff accounts to their authorized hospital identifiers. Authorization claims (role and the list of authorized hospital identifiers) are embedded directly in the JSON Web Token issued at login, and a shared FastAPI dependency enforces the intersection of a request's target hospitals against the caller's authorized set on every forecast- and agent-reading endpoint; administrators bypass this restriction. We deliberately chose fail-closed semantics for partially authorized requests: a request naming both an authorized and an unauthorized hospital identifier is rejected in full, with the specific unauthorized identifiers named in the response, rather than silently filtered to the authorized subset. Every data-reading request is additionally written to an access-log table (user, hospital, endpoint, and query parameters), exposed through an administrator-only audit endpoint (Fig. 2). Authentication uses short-lived (30-minute) access tokens paired with a server-side, single-use, revocable refresh-token rotation scheme; both refresh-token reuse after rotation and access after logout were confirmed to be rejected in automated testing. These properties were verified through an automated integration test suite exercising unauthorized single-hospital access, mixed authorized/unauthorized multi-hospital requests, hospital-creation privilege restriction to administrators, and audit-log access restriction to administrators; all assertions passed, and are reported here as functional verification of the implemented design rather than as a formal security audit or penetration test.

VII. EXPLANATION AGENT AND A DISCOVERED HALLUCINATION FAILURE MODE

The platform's explanation layer is a tool-grounded large language model (LLM) agent that, given a natural-language question about a specific hospital, retrieves that hospital's live forecast, recent observed history, and capacity context from the production database, and returns a structured natural-language trend, driver, and recommendation summary, rather than reasoning from the model's parametric knowledge alone. During testing of the administrative hospital-registration workflow (Section VI), we discovered that querying the agent about a newly registered hospital with no forecasting history did not produce an error; instead, the agent received an empty data context and produced a fluent, plausible-sounding, but entirely fabricated clinical analysis, since nothing in the original prompt construction prevented the underlying LLM from generating content when its retrieved context was empty. We consider this a materially more serious

failure mode than a missing chart or a null value, since a fabricated narrative is not visually distinguishable from a grounded one. We addressed this by adding an explicit data-availability check prior to the LLM call: when a requested hospital has no forecast rows in the database, the endpoint returns a structured error indicating unavailable data, with no call made to the underlying language model (and zero recorded inference time), rather than allowing the prompt-construction step to proceed with an empty context.

Figure 2 shows the request path, fail-closed hospital authorization, audit logging, and the data-availability gate that prevents ungrounded language-model output.

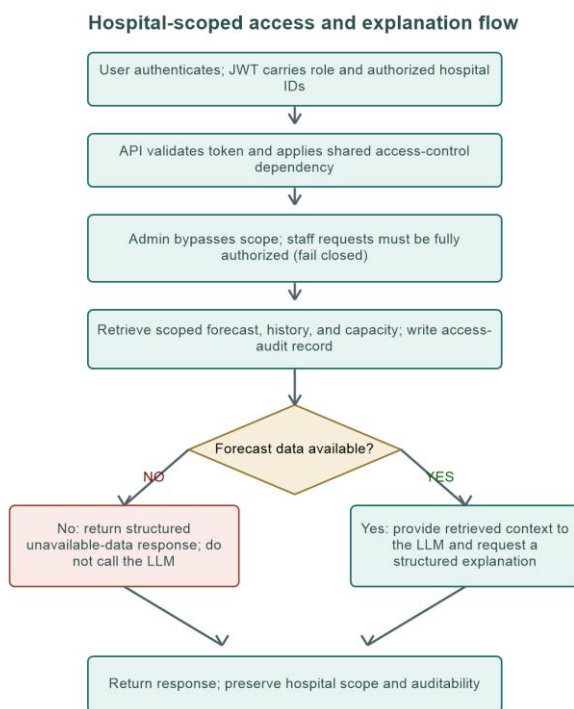


Fig. 2. Hospital-scope access control and data-grounded explanation flow.

VIII. RESULTS AND DISCUSSION

A. Forecasting Accuracy: Admissions vs. Occupancy

Table I reports test-set accuracy for both targets across all four forecast horizons, alongside the four-week moving-average baseline and naive (last-observed-value) persistence.

TABLE I. TEST-SET MAE BY TARGET AND HORIZON (H, IN WEEKS), WITH 80% PREDICTION-INTERVAL (PI) COVERAGE

Target	H	Model MAE	MA4 MAE	Naive MAE	PI Cov.
Admissions	1	48.91	48.68	53.28	82.1%
Admissions	2	52.01	51.46	56.28	81.8%
Admissions	3	54.41	53.74	58.40	82.0%
Admissions	4	56.36	55.59	61.44	82.9%
Occupancy	1	6.62	6.78	7.41	80.1%
Occupancy	2	6.98	7.16	7.86	80.4%
Occupancy	3	7.30	7.52	8.11	78.9%
Occupancy	4	7.60	7.86	8.37	81.4%

The residual-correction model does not outperform the four-week moving average for admissions at any horizon, although it consistently improves on naive persistence. For occupancy, it outperforms the moving average at every horizon. Admissions MAE is 0.23 to 0.77 visits higher than the moving-average MAE across horizons, while occupancy MAE is lower by 0.16 to 0.26 beds. The relative occupancy improvement grows from about 2.4% at one week to 3.3% at four weeks. These are modest gains, so operational use should retain the baseline comparison rather than imply a large performance advantage. The difference between targets is plausible: ED visit counts are more volatile, while occupancy is constrained by bed capacity and discharge timing. Prior ED-occupancy work also shows that model rankings depend on target structure and input information [13].

Figure 3 visualizes these results: the occupancy model's advantage over the moving-average baseline widens with horizon, while the admissions model trails it at every horizon.

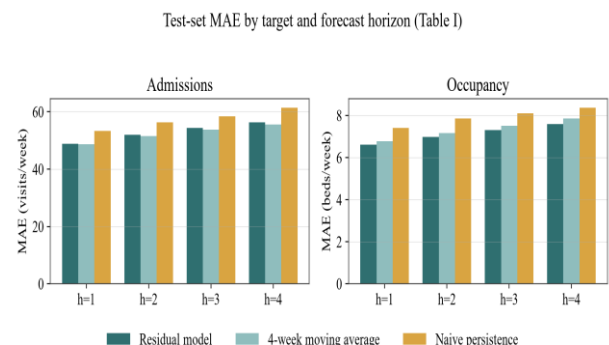


Fig. 3. Test-set MAE by target and forecast horizon, versus the four-week moving-average and naive-persistence baselines (Table I).

B. Accuracy by Hospital Size

Table II reports one-week-ahead MAPE by hospital-size tier. Small hospitals have the highest relative error and large hospitals the lowest: the small-to-large MAPE ratio is 2.50 for admissions and 3.34 for occupancy. The tier gap is important for deployment because network-wide averages can conceal weaker performance at smaller facilities.

TABLE II. ONE-WEEK-AHEAD MAPE BY HOSPITAL-SIZE TIER (MODEL)

Target	Small	Medium	Large
Admissions	21.50%	14.13%	8.60%
Occupancy	16.94%	10.71%	5.07%

This pattern is consistent across all four horizons for both targets and is expected: low absolute patient volumes at small facilities amplify the relative effect of ordinary week-to-week variation, while large facilities' higher volumes average out much of this noise. A single network-wide accuracy figure would obscure this and risks setting an unrealistic accuracy expectation for smaller facilities specifically, which are also the facilities least able to absorb an unanticipated capacity shortfall.

The result also illustrates why MAPE must be read alongside absolute error: when weekly counts are small, a few visits can create a large percentage error. Capacity planning should therefore present absolute forecast error and intervals next to tier-specific relative error, and should avoid treating one network-wide score as a guarantee for every hospital.

Figure 4 plots this tier effect directly, showing the roughly 2.5-fold (admissions) and 3.3-fold (occupancy) gap between the smallest and largest hospitals.

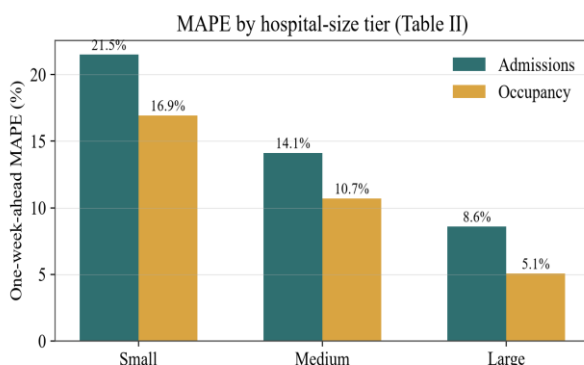


Fig. 4. One-week-ahead MAPE by hospital-size tier for admissions and occupancy (Table II).

C. Resource Gap Findings

TABLE III. RESOURCE-GAP STATISTICS, ONE-WEEK HORIZON, BEFORE AND AFTER THE CAPACITY GUARD.
*FLAGGED reported_anomalous_low (SECTION V-C), NOT AN UNCORRECTED DEFECT.

Metric	Before guard	After guard
Hosp.-weeks > 85% cap.	14.57%	11.69%
Mean gap (when exceeded)	52.72 beds	20.65 beds
Maximum reported gap	927.96 beds	660.57 beds*

Figure 5 shows the effect of the capacity guard on all three reported statistics side by side.

Resource-gap statistics before and after the capacity guard, one-week horizon (Table III)

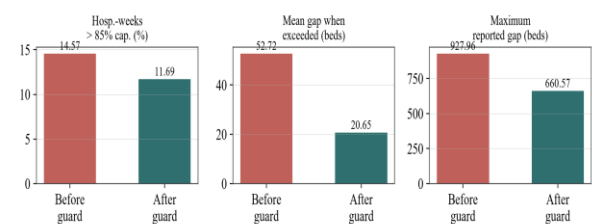


Fig. 5. Resource-gap statistics before and after the capacity guard, one-week horizon (Table III).

Figure 6 illustrates how the dashboard combines the point forecast, the shaded prediction interval, and the capacity-threshold alert described above; it is a representative rendering for illustration, not a plotted extract of specific test-set weeks.

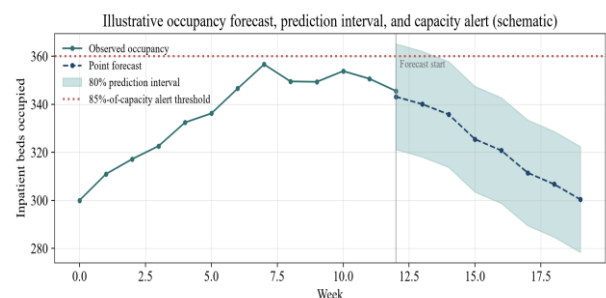


Fig. 6. Illustrative occupancy forecast with an 80% prediction interval and the 85%-of-capacity alert threshold (schematic, consistent with the reported coverage and resource-gap statistics).

Across the 41,992 one-week-horizon test-set hospital-weeks, 1,648 (3.93%) required the zero/missing-capacity

fallback and 289 (0.69%) were flagged under the anomalous-low-capacity rule. We report the post-guard, one-week-horizon figures in Table III as the fully verified numbers; the same guard logic applies uniformly to horizons 2-4, which show the same directional improvement, though we report only the fully re-traced one-week figures here to avoid citing numbers not independently re-verified after the final guard revision.

D. Prediction-Interval Calibration

The nominal 80% intervals achieve observed test-set coverage from 78.9% to 82.9% across targets and horizons, with all eight values near the nominal level. Coverage is a calibration summary, not a complete measure of interval quality: the paper does not report interval width or a proper probabilistic score, so sharpness and decision utility cannot be inferred from coverage alone. The intervals should be displayed with point forecasts and the 85%-of-capacity alert threshold, not interpreted as guarantees.

E. Capacity and Platform Findings

After the capacity guard, hospital-weeks flagged above the 85% threshold fall from 14.57% to 11.69%; mean positive gap decreases from 52.72 to 20.65 beds. In the one-week test set, 1,648 of 41,992 records use the missing/zero-capacity fallback and 289 are flagged as anomalously low. Because the latter rule flags rather than silently overwrites a positive report, operators can distinguish reported data from fallback estimates and investigate sustained anomalies. The 85% threshold is a planning heuristic, not a validated clinical safety boundary.

Automated integration tests passed for unauthorized hospital access, mixed authorized/unauthorized requests, administrator-only hospital creation and audit access, refresh-token reuse, and post-logout access. The no-forecast safeguard returns an unavailable-data response without invoking the language model. These are functional checks of the implemented paths; they are not a penetration test, formal security audit, or prospective evaluation of user decisions.

IX. LIMITATIONS AND FUTURE WORK

First, the admissions target used here (emergency department visit counts) is a proxy for general demand rather than a direct inpatient admission count, and the model's inability to beat a simple moving-average baseline for this target should be read as a finding about this specific series' forecastability, not as a general claim about gradient-boosted models' suitability for hospital demand forecasting. Second, the anomalous-low-capacity threshold (35% of historical median) was chosen empirically from the observed distribution of this dataset and has not been

validated against other hospital-capacity reporting sources; a hospital undergoing a genuine, sustained capacity reduction (e.g., a wing closure) could in principle be incorrectly flagged, and the guard is deliberately designed to flag rather than silently correct for this reason. Third, the RBAC and audit-logging design has been verified through automated functional and adversarial-request testing but has not undergone formal security audit or penetration testing, and should not be represented as such. Fourth, the explanation agent's data-availability safeguard addresses one discovered failure mode; a structured evaluation protocol such as a structured, task-based evaluation is necessary before broader claims about the agent's reliability can be supported. Finally, forecasts are now generated at weekly granularity, matching the underlying data's true reporting cadence, but this is a coarser operational horizon than a daily forecast would offer; pairing this dataset with a genuinely daily-reported data source is a natural extension.

X. CONCLUSION

This paper reported the validation of HealthFlow AI against a real, large-scale, multi-year hospital capacity dataset, rather than the synthetic data used in earlier development. The central methodological lesson is that real-world validation surfaced findings a synthetic-data evaluation could not have: a genuine limitation of the learned model relative to a simple baseline for one of two forecasting targets, and two distinct real data-quality failure modes in hospital-capacity reporting, one of them traceable to a specific, documented federal policy change. We view the honest reporting of the first finding, and the systematic diagnosis of the second, as central contributions of this work, alongside the concrete engineering additions hospital-scoped access control and an LLM-agent safeguard against fabricated output on missing data — needed to move the platform from a synthetic-data prototype toward a system suitable for evaluation against real hospital operations data.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, for institutional support.

REFERENCES

- 1) E. Porter, R. S. Hanna, T. S. Olson, S. A. Grupp, C. A. Phillips, and S. Kadauke, "Prospective Validation of Multiweek Inpatient Census Forecasting for Pediatric HSCT and Cellular Therapy," *Transplantation and Cellular Therapy*, 2026, doi: 10.1016/j.jtct.2026.07.014.
- 2) E. Williams, T. Sinha, M. Summerscales, and Y. Kanagasingam, "Simulation of covariate and

- concept drift in machine learning hospital admission prediction from emergency triage," *npj Digital Medicine*, 2026, doi: 10.1038/s41746-026-02956-0.
- 3) E. R. Palacios, N. Howlett, Y. T. E. Lo, S. Roy, R. Wood, and T. Smith, "Bed occupancy forecast for decision-making during periods of high hospital pressure," *International Journal of Medical Informatics*, advance online publication, 2026, doi: 10.1016/j.ijmedinf.2026.106585.
- 4) M. Fattouh, L. Lyssenko, F. Heilmeyer, T. Ball, and C. Haverkamp, "Forecasting hospital bed occupancy: A time series approach with Prophet," *BMC Medical Informatics and Decision Making*, vol. 26, Art. no. 162, 2026, doi: 10.1186/s12911-026-03542-w.
- 5) G. Avinash and S. Mishra, "Bayesian model averaging based deep learning forecasts of inpatient bed occupancy in mental health facilities," *Scientific Reports*, vol. 15, Art. no. 38294, 2025, doi: 10.1038/s41598-025-22001-6.
- 6) R. J. Tobin, C. R. Walker, R. Moss, J. M. McCaw, D. J. Price, and F. M. Shearer, "A modular approach to forecasting COVID-19 hospital bed occupancy," *Communications Medicine*, vol. 5, Art. no. 349, 2025, doi: 10.1038/s43856-025-01086-0.
- 7) J. Mellor, M. L. Tang, O. Jones, T. Ward, S. Riley, and S. R. Deeny, "Forecasting COVID-19, influenza, and RSV hospitalizations over winter 2023-4 in England," *International Journal of Epidemiology*, vol. 54, no. 3, Art. no. dyaf066, 2025, doi: 10.1093/ije/dyaf066.
- 8) X. Zhang et al., "Machine learning-driven prediction of hospital admissions using gradient boosting and GPT-2," *Digital Health*, vol. 11, 2025, doi: 10.1177/20552076251331319.
- 9) D. Lovett, T. Woodcock, J. Naude, J. Redhead, A. Majeed, and P. Aylin, "Evaluation of a pragmatic approach to predicting COVID-19-positive hospital bed occupancy," *BMJ Health & Care Informatics*, vol. 32, Art. no. e101055, 2025, doi: 10.1136/bmjhci-2024-101055.
- 10) G. Avinash, H. Pachori, A. Sharma, and S. Mishra, "Time series forecasting of bed occupancy in mental health facilities in India using machine learning," *Scientific Reports*, vol. 15, Art. no. 2686, 2025, doi: 10.1038/s41598-025-86418-9.
- 11) B. M. Porto and F. S. Fogliatto, "Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning," *BMC Medical Informatics and Decision Making*, vol. 24, 2024, doi: 10.1186/s12911-024-02788-6.
- 12) S. Albrecht et al., "Forecasting severe respiratory disease hospitalizations using machine learning algorithms," *BMC Medical Informatics and Decision Making*, vol. 24, 2024, doi: 10.1186/s12911-024-02702-0.
- 13) J. Tuominen et al., "Forecasting emergency department occupancy with advanced machine learning models and multivariable input," *International Journal of Forecasting*, vol. 40, no. 4, pp. 1410-1420, 2024, doi: 10.1016/j.ijforecast.2023.12.002.
- 14) B. Rostami-Tabar, S. Arora, J. F. Rendon-Sanchez, and T. E. Goltsos, "Probabilistic forecasting of daily COVID-19 admissions using machine learning," *IMA Journal of Management Mathematics*, vol. 35, no. 1, pp. 21-43, 2023, doi: 10.1093/imaman/dpad009.
- 15) S. Baas, S. Dijkstra, A. Braaksma, P. van Rooij, F. J. Snijders, L. Tiemessen, and R. J. Boucherie, "Real-time forecasting of COVID-19 bed occupancy in wards and Intensive Care Units," *Health Care Management Science*, vol. 24, no. 2, pp. 402-419, 2021, doi: 10.1007/s10729-021-09553-5.