

Machine Learning-Assisted Demand Forecasting and Inventory Optimization of Novel Drug Delivery Systems in Tertiary Care Hospitals: A Case Study from Amaravati, Andhra Pradesh

¹Dr. Nedunuri Manikya Jyothirmmai, ²Dr. K J Sumitha, ³Ms. B. Katharine

¹Assistant Professor, School of Management Studies

JNTUK Kakinada, Andhra Pradesh, India

Email id : mjnedunuri@gmail.com

²Assistant Professor, AMET Business School, Academy of Maritime Education and Training (Deemed to be University), Chennai, India

Email id : kjsmitha@gmail.com

³Student, BE (Marine Engineering), 3rd Year, AMET Business School
Academy of Maritime Education and Training (Deemed to be University), Chennai, India.

Email id : kancherlahope5@gmail.com

Abstract

Purpose: This study develops, validates, and evaluates a machine learning (ML) framework for demand forecasting and inventory optimization of novel drug delivery systems (NDDS) in tertiary care hospitals, using the Amaravati capital region of Andhra Pradesh as the case context.

Design/methodology/approach: A convergent mixed-methods design was employed. Qualitative data comprised 12 semi-structured interviews with hospital pharmacists, procurement officers, clinical pharmacologists, and administrators, analysed thematically and reported per COREQ. Quantitative data comprised 36 months of dispensing records for 24 NDDS formulations across three tertiary care hospitals (n = 864 formulation-months). Formulations were classified using the Syntetos–Boylan demand categorization framework (ADI and CV²). Four forecasting models (XGBoost, LSTM, Prophet, SARIMAX) were benchmarked against the incumbent moving-average policy using rolling-origin validation with 24 forecast origins. Inventory policies were calibrated using a continuous-review (Q, R) framework with EOQ and safety-stock logic at a 95% service level. Sensitivity analysis examined holding-cost rate, lead time, service level, and baseline wastage assumptions.

Findings: Demand-pattern classification revealed that only 29.2% of NDDS formulations exhibited smooth demand; 29.2% were intermittent, 20.8% lumpy, and 20.8% erratic. A class-matched hybrid model (XGBoost for smooth/erratic; LSTM for intermittent/lumpy) achieved MAPE 7.1% (95% CI: 6.3–7.9), RMSE 12.9 units, and bias +0.3 units, against a baseline MAPE of 18.5% (95% CI: 16.8–20.2). LSTM significantly outperformed XGBoost within intermittent (7.4% vs 10.4%, p = 0.018) and lumpy (12.4% vs 13.8%, p = 0.036) classes, while XGBoost led within smooth (4.2% vs 5.1%, p = 0.022) and erratic (5.6% vs 6.3%, p = 0.041) classes. ML-assisted ordering reduced average inventory holding cost from ₹9.62 to ₹6.58 lakh/month (–31.5%), near-expiry wastage from 5.9% to 2.0% of procured value (–66.1%), stockout episodes from 47 to 9 formulation-months (–80.9%), and raised service level from 83.1% to 96.4%. Net annual saving was ₹1.62 crore against annualized implementation cost of ₹23 lakh (benefit–cost ratio 7.0:1; payback 2.7 months).

Originality/value: This is the first study to (a) disaggregate NDDS from conventional pharmaceuticals in ML forecasting, (b) apply demand-pattern classification to model selection for hospital drug forecasting in India, and (c) quantify the full cost–availability trade-off of ML-assisted NDDS inventory management in a Tier-II Indian capital region.

Keywords: machine learning; demand forecasting; inventory optimization; novel drug delivery systems; hospital pharmacy; Syntetos–Boylan classification; India

How to cite this article: 1Dr. Nedunuri Manikyam Jyothirmmai 2Dr. K J Sumitha 3Ms. B. Katharine. Machine Learning-Assisted Demand Forecasting and Inventory Optimization of Novel Drug Delivery Systems in Tertiary Care Hospitals: A Case Study from Amaravati, Andhra Pradesh. *Int J Drug Deliv Technol*. 2026;16(83s):174–185. DOI: 10.25258/ijddt.16.83s.20.

1. Introduction

1.1 Background

Hospital pharmaceutical inventory management faces a structural tension: ensuring uninterrupted availability of clinically critical medicines while avoiding the financial and clinical costs of

overstocking, expiry, and obsolescence (Volland et al., 2017). This tension is amplified for novel drug delivery systems (NDDS)—formulations engineered for controlled release, targeted delivery, or enhanced bioavailability, including liposomal, nanoparticle-based, long-acting injectable, transdermal, and

implantable systems. NDDS formulations are typically higher-cost, have shorter or more storage-sensitive shelf lives, and exhibit more variable demand than conventional generics, rendering them disproportionately vulnerable to both stockouts and wastage (Perlman & Levner, 2014).

Indian public hospital supply chains compound these difficulties. Andhra Pradesh's drug supply system, administered through APMSIDC via the e-Aushadhi platform, has documented structural weaknesses: as of early 2026, 17 medicines were entirely unavailable across the state, and 178 formulations had less than one month of stock cover. The state government has announced reforms including automated alert systems, real-time dashboards, and prescription auditing, implicitly acknowledging that the existing procurement-to-dispensing chain is not demand-responsive.

1.2 The Amaravati Context

While Amaravati serves as an ideal context for rapidly expanding Tier-II health systems, operational policies must be recalibrated prior to deployment in different regional infrastructure profiles. In dense Tier-I metropolitan hospitals with hyper-reliable vendor logistics and automated health information systems, safety stock thresholds (σ -scores) can be safely lowered to maximize capital efficiency. Conversely, for Tier-III or rural facilities facing volatile supplier lead times, unstable cold-chain transit, and power grid intermittency, the inventory policy must incorporate expanded lead-time variance (σ_{LT}) and adjusted safety stock buffers. Systematically tuning the continuous-review (Q, R) parameters to match regional supply-chain maturity ensures system resilience across diverse geographic operational realities.

Amaravati, the capital region of Andhra Pradesh, presents an unusually instructive case. The city is undergoing rapid tertiary care expansion: a 600-bed, 32-specialty facility; a 500-bed multi-specialty hospital under a long-term lease arrangement; a ₹700-crore medical college with a 500-bed teaching hospital; and an announced ~1,000-bed quaternary care facility. These institutions will collectively manage a substantial and growing NDDS portfolio. Yet no published evidence exists on how their inventory systems should be adapted for these formulations, and none of the three hospitals in this study operated dedicated demand-forecasting software at the time of data collection.

1.3 Research Gap

Three specific gaps motivate this study.

First, no published ML forecasting study has isolated NDDS from conventional pharmaceuticals. Existing hospital pharmacy forecasting research treats drug demand as a homogeneous category or stratifies only

by therapeutic class, despite NDDS having fundamentally different cost, shelf-life, storage, and demand-variability profiles (Ahmadi et al., 2019).

Second, Indian hospital pharmacy analytics research is concentrated in metropolitan centres. Tier-II capital regions—which are absorbing the fastest growth in tertiary care capacity—remain unexamined, despite their distinct supply-chain characteristics: longer vendor lead times, thinner cold-chain infrastructure, and less mature health information systems.

Third, existing studies predominantly report forecast accuracy improvements without quantifying the operational trade-off. A model that reduces MAPE may nonetheless increase ordering frequency, strain cold-chain handling capacity, or fail to reduce wastage. The forecast-to-inventory-to-clinical-outcome pathway is rarely made explicit.

1.4 Research Questions

- **RQ1:** What demand patterns characterise NDDS in Amaravati tertiary care hospitals, and how do these differ from conventional pharmaceuticals?
- **RQ2:** Which ML model architecture minimises forecast error for NDDS demand, and does the optimal architecture vary systematically by demand-pattern class?
- **RQ3:** What is the quantifiable inventory cost, stockout, wastage, and clinical-availability impact of ML-assisted ordering compared with incumbent practice?

1.5 Contribution

The study makes three contributions. Methodologically, it demonstrates that demand-pattern classification should precede model selection in hospital drug forecasting—a step largely absent from the hospital pharmacy ML literature. Empirically, it provides the first NDDS-specific forecasting and inventory optimization evidence from an Indian Tier-II capital region. Practically, it quantifies the full cost–benefit case, including implementation cost, sensitivity to assumptions, and non-financial clinical availability outcomes, thereby offering hospital administrators a decision-ready framework.

2. Literature Review

2.1 Forecasting Methods for Intermittent and Non-Stationary Demand

The theoretical foundation for demand-pattern-aware forecasting rests on the Syntetos–Boylan framework (Syntetos et al., 2005), which classifies demand series along two dimensions: average demand interval (ADI) and the squared coefficient of variation of demand sizes (CV^2). The framework yields four archetypes—smooth, erratic, intermittent, and lumpy—with established cut-offs ($ADI = 1.32$; $CV^2 = 0.49$). Syntetos and Boylan (2005) demonstrated that forecast

accuracy and the relative ranking of methods vary systematically across these classes, implying that no single model is universally optimal.

This has direct implications for ML application. Gradient-boosted trees such as XGBoost (Chen & Guestrin, 2016) excel at capturing non-linear interactions among dense, feature-rich series, and have shown stable error across therapeutic groups in recent hospital pharmacy applications. Conversely, long short-term memory networks (Hochreiter & Schmidhuber, 1997) can learn dependencies across variable-length gaps, which is theoretically advantageous for sparse series where zero-demand periods carry information. Prophet (Taylor & Letham, 2018) handles multiple seasonality and change points with minimal tuning but assumes a largely additive structure, while SARIMAX (Box et al., 2015) requires stationarity-inducing transformations that are awkward for intermittent series.

Benchmarking evidence from the M4 competition (Makridakis et al., 2020) confirms that hybrid and combination approaches generally outperform single models, and Bergmeir and Benítez (2012) established that rolling-origin cross-validation is the appropriate evaluation scheme for time-series predictors, as standard k-fold cross-validation leaks future information.

2.2 Hospital Inventory Optimization

The classical continuous-review (Q, R) policy—with order quantity from the economic order quantity model and reorder point derived from lead-time demand plus safety stock—remains the workhorse of hospital pharmaceutical inventory control (Silver et al., 2016). Its applicability to perishable and specialised pharmaceuticals has been examined by Perlman and Levner (2014), who showed that expiry costs fundamentally alter the optimal order quantity downward relative to the classical EOQ solution. Volland et al. (2017) and Ahmadi et al. (2019), in comprehensive reviews of hospital material logistics, both identify the absence of demand-analytics integration as a persistent gap.

2.3 NDDS-Specific Inventory Considerations

NDDS formulations introduce four distinct inventory complexities: (i) higher unit cost, which raises the financial consequence of both overstocking and stockout; (ii) storage constraints, particularly cold-chain requirements for liposomal and nanoparticle products; (iii) regulatory constraints—in India, transdermal fentanyl and buprenorphine fall under the Narcotic Drugs and Psychotropic Substances Act, 1985, which imposes separate licensing and stocking limits; and (iv) limited therapeutic substitutability, meaning a stockout cannot always be resolved by dispensing an alternative. These factors collectively

imply that NDDS inventory policies should not be inherited from conventional drug categories.

2.4 Andhra Pradesh Supply Chain Context

Andhra Pradesh's drug procurement operates through APMSIDC and the e-Aushadhi platform, which logs transactions but provides no analytical layer. The state's planned reforms—including real-time stock dashboards spanning primary to tertiary facilities—signal political recognition of the need for supply-chain modernisation, but the analytical capability required to convert transactional data into forward-looking demand signals remains absent.

3. Methodology

3.1 Study Design

A convergent mixed-methods design was employed, with qualitative and quantitative strands collected and analysed in parallel and integrated at the interpretation stage. The study was conducted across three tertiary care hospitals in the Amaravati capital region: H1, a 600-bed super-specialty hospital; H2, a 500-bed multi-specialty hospital; and H3, a 350-bed teaching hospital. Data spanned January 2023 to December 2025.

3.2 Qualitative Component

3.2.1 Sampling and Participants

Purposive sampling was used to recruit 12 key informants: five hospital pharmacists, four procurement officers, two clinical pharmacologists, and one hospital administrator. Sample size was determined by thematic saturation, achieved after 10 interviews; two additional interviews confirmed no new themes emerged. Participant characteristics are reported in Appendix B.

3.2.2 Data Collection

Semi-structured interviews (45–70 minutes) followed an interview guide (Appendix C) exploring current NDDS procurement practice, demand-estimation methods, stockout and expiry experience, and perceived barriers to analytics adoption. Interviews were audio-recorded, transcribed verbatim, and returned to participants for correction.

3.2.3 Analysis and Rigour

Thematic analysis followed Braun and Clarke (2006). Two researchers independently coded all transcripts; inter-coder reliability was substantial (Cohen's $\kappa = 0.82$; Cohen, 1960). Discrepancies were resolved by consensus. Reporting adhered to the Consolidated Criteria for Reporting Qualitative Research (COREQ; Tong et al., 2007); the completed checklist is provided as Appendix D. Member checking was conducted with four participants, all of whom confirmed thematic accuracy.

3.3 Quantitative Component

3.3.1 Data Source and Preparation

De-identified dispensing and procurement records for 24 NDDS formulations were extracted from hospital

pharmacy management systems and cross-validated against e-Aushadhi transaction logs. Data quality assessment identified 8.4% of records with missing unit-cost fields (imputed from the corresponding purchase order) and 3.1% duplicate entries (removed). The final dataset comprised 864 formulation-month observations.

3.3.2 Demand-Pattern Classification

Prior to forecasting, each formulation was classified using the Syntetos–Boylan framework. ADI was computed as the mean interval between non-zero demand periods; CV^2 as the squared coefficient of variation of non-zero demand sizes. Cut-offs were $ADI = 1.32$ and $CV^2 = 0.49$, yielding smooth, erratic, intermittent, and lumpy classes (Syntetos et al., 2005).

3.3.3 Forecasting Models

All models were implemented in Python 3.11 (scikit-learn 1.4, TensorFlow 2.15, Prophet 1.1, statsmodels 0.14) with Optuna-based Bayesian hyperparameter optimization over a 5-fold rolling-origin inner loop.

XGBoost. Gradient-boosted trees (Chen & Guestrin, 2016) with features: lag-1, lag-2, lag-3 demand; rolling 3- and 6-month means; month-of-year dummies; clinician-initiated order flag; count of active prescribers; lagged procurement lead time; festival/holiday indicator; prior-stockout indicator; and ward bed-occupancy rate. Best parameters: $n_estimators = 500$, $max_depth = 6$, $learning_rate = 0.05$, $subsample = 0.8$, $colsample_bytree = 0.8$, $min_child_weight = 3$.

LSTM. Two-layer architecture (64 and 32 units), dropout 0.2, lookback window 6 months, Adam optimizer (learning rate 0.001), 100 epochs with early stopping (patience 10). Best configuration retained a single dropout layer and tanh activation.

Prophet. Additive trend with yearly seasonality, automatic changepoint detection (prior scale 0.05), and Indian public-holiday effects.

SARIMAX. Seasonal ARIMA with exogenous regressors (lead time, lagged stockout indicator), orders selected by AIC over a constrained grid ($p, d, q \in [0, 2]$; $P, D, Q \in [0, 1]$; $m = 12$).

Baseline. The hospitals' incumbent 3-month moving-average policy.

3.3.4 Evaluation Protocol

Rolling-origin validation with 24 forecast origins (expanding window), one-month-ahead horizon. Metrics: MAPE, RMSE, and forecast bias (mean error). Ninety-five percent confidence intervals were computed across origins using the bootstrap (10,000 resamples). Pairwise comparisons used the Wilcoxon signed-rank test on absolute percentage errors ($\alpha = 0.05$).

3.3.5 Model Interpretation

SHAP values (Lundberg & Lee, 2017) were computed for the XGBoost model; permutation importance was

used for the LSTM, as SHAP for recurrent architectures is computationally prohibitive at this scale.

3.3.6 Inventory Optimization

A continuous-review (Q, R) policy was calibrated per formulation:

- **Q** from the EOQ formula, modified for perishability: $Q^* = \min(Q_EOQ, Q_perishability)$, where $Q_perishability$ is the quantity whose expected cycle time does not exceed 70% of remaining shelf life (Perlman & Levner, 2014).
- **R** = forecast demand during lead time + safety stock, with safety stock = $z \times \sigma_LT$ at a 95% service level ($z = 1.645$).
- Holding cost rate = 22.5% of unit value annually, decomposed as 8% capital, 6% storage and cold chain, 5% insurance and security, 3.5% obsolescence risk.

Backtesting simulated 12 months of operation (January–December 2025) under both ML-assisted and historical ordering policies, holding demand realisations constant.

3.3.7 Sensitivity Analysis

A one-at-a-time sensitivity analysis (tornado) varied four parameters across plausible ranges: holding-cost rate (15–30%), procurement lead time (± 2 weeks), service level (90–99%), and baseline wastage rate (4–8%). Net annual saving was recomputed for each scenario.

4. Results

4.1 Qualitative Findings

Three themes emerged.

Theme 1: Reactive, request-driven procurement. All procurement officers described ordering as triggered by clinical requests and vendor availability rather than systematic estimation. One pharmacist stated: *"For NDDS, we order when the consultant asks. There is no forecast. If we run out, we borrow from another hospital or do an emergency purchase at higher cost."*

Theme 2: Transactional, not analytical, digital infrastructure. Participants consistently distinguished between record-keeping and forecasting. *"It tells us what we consumed, not what we will need,"* observed one procurement officer. None of the three hospitals operated forecasting software.

Theme 3: Rational risk-averse understocking. Pharmacists expressed acute expiry anxiety for high-cost NDDS. *"These are expensive medicines. If we stock three months and a patient does not come, it expires. So we stock one month, and then we run out,"* explained a clinical pharmacologist. This behaviour was corroborated quantitatively: average

NDDS stock cover was 2.17 months, against 3.4 months for conventional drugs.

4.2 Quantitative Findings

4.2.1 Descriptive Statistics and Demand Classification (RQ1)

Table 1 presents descriptive statistics for 12 representative formulations; the full 24-formulation table is provided in Appendix A.

Table 1. Descriptive statistics and Syntetos–Boylan classification of representative NDDS formulations (36 months)

Sl. No.	Formulation	Class	Mean monthly demand (units)	SD	ADI	CV ²	Unit cost (₹)	Shelf life (mo)	Storage
1	Pegylated liposomal doxorubicin 2 mg/mL	Smooth	18.4	5.7	1.04	0.31	62,400	18	2–8 °C
2	Liposomal amphotericin B 50 mg	Lumpy	6.2	5.1	2.48	0.83	8,950	24	2–8 °C
3	Albumin-bound paclitaxel 100 mg	Erratic	11.8	8.4	1.11	0.58	14,200	24	RT

4	Paliperidone palmitate LAI 150 mg	Intermittent	4.7	1.9	1.94	0.34	21,600	36	RT
5	Aripiprazole lauroxil LAI 441 mg	Lumpy	2.3	1.7	2.35	0.71	26,800	24	RT
6	Fentanyl transdermal 25 µg/h	Erratic	42.6	27.3	1.18	0.62	1,180	24	RT (NDPS)
7	Buprenorphine transdermal 10 µg/h	Intermittent	9.1	3.4	1.62	0.41	2,340	24	RT (NDPS)
8	Goserelin implant 3.6 mg	Intermittent	7.8	2.6	1.55	0.29	9,800	36	RT
9	Leuprolide implant 11.25 mg	Lumpy	4.1	3.0	2.10	0.66	12,400	24	RT
10	Nanosomal docetaxel	Lumpy	5.4	4.6	2.62	0.91	18,7	18	2–8 °C

	20 mg/mL						00		
11	Insulin glargine 100 IU/mL	Smooth	268.4	61.2	1.022	0.222	890	24	2–8 °C
12	Pegfilgrastim 6 mg	Erratic	14.2	10.9	1.214	0.74	11300	24	2–8 °C

Class distribution across all 24 formulations: smooth n = 7 (29.2%); erratic n = 5 (20.8%); intermittent n = 7 (29.2%); lumpy n = 5 (20.8%). Notably, **70.8% of NDDS formulations exhibited non-smooth demand**, against 41.2% among a comparison set of 68 conventional formulations in the same hospitals ($\chi^2 = 6.84$, $p = 0.009$). Cold-chain-requiring formulations constituted 45.8% of the NDDS portfolio.

4.2.2 Forecast Accuracy (RQ2)

Table 2 reports accuracy overall and by demand class. **Table 2. Forecast accuracy by model and demand class (MAPE %, rolling-origin validation, 24 origins)**

Model	Smooth (n=7)	Erratic (n=5)	Intermittent (n=7)	Lumpy (n=5)	Overall MAPE (95% CI)	RMSE	Bias
Moving average (baseline)	12.6	15.9	21.3	25.2	18.5 (16.8–20.2)	31.2	+4.7
SARI MAX	8.4	10.8	14.6	18.1	12.7 (11.4–	23.4	+2.9

					14.0)		
Prophet	6.8	8.9	11.9	16.5	10.8 (9.6–12.0)	19.8	+1.8
LSTM (uniform)	5.1	6.3	7.4	12.4	7.5 (6.7–8.3)	14.2	–0.9
XGBoost (uniform)	4.2	5.6	10.4	13.8	8.3 (7.4–9.2)	15.1	+1.1
Hybrid (class-matched)	4.2	5.6	7.4	12.4	7.1 (6.3–7.9)	12.9	+0.3

The class-matched hybrid model reduced MAPE by 61.6% relative to baseline (Wilcoxon $p < 0.001$) and by 44.1% relative to the best single uniform model. Critically, the two leading models were statistically indistinguishable *on average* (LSTM 7.5% vs XGBoost 8.3%; $p = 0.087$) but differed *significantly within demand classes*:

- XGBoost superior for smooth (4.2% vs 5.1%; $p = 0.022$) and erratic (5.6% vs 6.3%; $p = 0.041$)
- LSTM superior for intermittent (7.4% vs 10.4%; $p = 0.018$) and lumpy (12.4% vs 13.8%; $p = 0.036$)

This is the study's principal methodological finding: **aggregate model comparison masks class-specific superiority**, and model selection should therefore follow demand classification.

Forecast bias was negligible for the hybrid model (+0.3 units), indicating balanced over- and under-forecasting—an operationally important property, since systematic positive bias inflates expiry risk and negative bias inflates stockout risk.

4.2.3 Feature Importance

Table 3 reports the ten highest-ranked features by mean absolute SHAP value for the XGBoost model.

Table 3. SHAP feature importance, XGBoost (mean |SHAP| in demand units)

Rank	Feature	Mean SHAP	Interpretation
1	Lag-1 demand	4.82	Strong short-memory persistence
2	Rolling 3-month mean	3.91	Medium-term trend capture
3	Clinician-initiated order flag	3.24	Prescriber behaviour dominates over epidemiological signal
4	Lag-2 demand	2.76	Second-order persistence
5	Month-of-year	2.15	Oncology and psychiatric treatment cycles
6	Active prescriber count	1.88	New-adopter diffusion effect
7	Lagged procurement lead time	1.42	Availability-induced demand (supply creates demand)
8	Festival/holiday indicator	0.94	Care-seeking and scheduling effects
9	Prior-stockout indicator	0.81	Stockout carries forward into subsequent periods
10	Bed-occupancy rate	0.67	Weak but non-trivial capacity signal

Two findings merit emphasis. First, **clinician-initiated ordering behaviour ranked third**, ahead of seasonality and occupancy—confirming the qualitative theme that NDDS demand is substantially prescriber-driven rather than epidemiologically driven. Second, **lagged lead time ranked seventh**

and prior stockout ninth, indicating that historical supply constraints have shaped recorded demand, meaning that models trained on such data partially learn the supply chain rather than true patient need. This constitutes a systematic demand-censoring problem and is a material limitation addressed in Section 5.4.

For the LSTM (permutation importance), the top five features were lag-1 demand, inter-demand interval (recency), rolling 6-month mean, active prescriber count, and a seasonality index—notably, **inter-demand interval appeared second**, consistent with the model's advantage on sparse series where the timing of the last demand is highly informative.

4.2.4 Inventory Optimization Outcomes (RQ3)

Table 4 reports simulated inventory performance over the 12-month backtesting period.

Table 4. Inventory performance: historical practice versus ML-assisted policy (12-month backtest, 24 formulations)

Metric	Historical practice	ML-assisted policy	Change
Average inventory value (₹ lakh)	513	351	−31.6%
Average stock cover (months)	2.17	1.48	−31.8%
Average holding cost (₹ lakh/month)	9.62	6.58	−31.5%
Stockout episodes (formulation-months)	47	9	−80.9%
Service level (% demand met from stock)	83.1%	96.4%	+13.3 pp
Near-expiry wastage (% of procured value)	5.9%	2.0%	−66.1%
Emergency procurement orders (per year)	31	6	−80.6%
Inter-hospital borrowings (per year)	22	3	−86.4%

Order frequency (orders/month)	14.2	18.7	+31.7%
Mean cold-chain handling events/month	96	134	+39.6%

The stockout reduction from 47 to 9 formulation-months represents a service-level improvement achieved *simultaneously* with a 31.6% reduction in average inventory value—a Pareto improvement that would be unattainable under the incumbent policy without improved forecast accuracy. This confirms that the mechanism is forecast quality rather than inventory buffering.

However, the 39.6% increase in cold-chain handling events is a material operational side effect and is discussed in Section 5.3.

4.2.5 Cost Breakdown

Table 5. Annual cost impact of ML-assisted NDDS inventory management (₹ lakh, three hospitals combined)

Cost category	Direction	Annual amount (₹ lakh)
Reduced inventory holding cost	Saving	36.4
Reduced near-expiry wastage	Saving	110.8
Reduced emergency procurement premium	Saving	39.9
Gross savings		187.1
Increased ordering cost	Cost	(2.6)
Implementation cost (amortized, 3-yr)	Cost	(12.0)
Recurring analyst and maintenance cost	Cost	(11.0)
Total implementation cost		(25.6)
Net annual saving		161.5

Annual NDDS procurement value across the three hospitals was ₹2,840 lakh. Implementation comprised one-time costs of ₹36 lakh (software ₹18 lakh; HIS integration ₹12 lakh; training ₹6 lakh) amortized over three years, plus recurring costs of ₹11 lakh per year (1.0 FTE data analyst ₹6.5 lakh; software maintenance ₹4.5 lakh).

Benefit–cost ratio: 7.0:1. Payback period: 2.7 months.

The dominant saving category is **reduced expiry wastage (₹110.8 lakh, 59.2% of gross savings)**, not reduced holding cost (19.5%). This has an important managerial implication: the primary financial case for ML-assisted NDDS inventory management rests on wastage avoidance, not inventory carrying-cost reduction. Hospitals with lower baseline wastage rates will realise proportionally smaller benefits.

4.2.6 Sensitivity Analysis

Table 6. One-at-a-time sensitivity analysis of net annual saving (₹ lakh)

Parameter	Low scenario	Base	High scenario	Swing
Holding-cost rate (15% / 22.5% / 30%)	149.4	161.5	173.7	±7.6%
Lead time (–2 wk / base / +2 wk)	170.8	161.5	154.4	±5.1%
Service level (90% / 95% / 99%)	168.5	161.5	148.5	±6.2%
Baseline wastage rate (4% / 5.9% / 8%)	107.5	161.5	221.1	±35.2%

Net savings remained positive across all 12 scenarios tested, ranging from ₹107.5 lakh to ₹221.1 lakh. The result is **most sensitive to the assumed baseline wastage rate** (±35.2%) and least sensitive to procurement lead time (±5.1%). Crucially, however, the *ranking* of interventions—wastage reduction first, then emergency procurement, then holding cost—was stable across all scenarios.

5. Discussion

5.1 Principal Findings and Comparison with Indian Evidence

The hybrid model's MAPE of 7.1% compares favourably with reported accuracy in Indian hospital pharmaceutical forecasting studies, which typically range from 8% to 15%. The 31.5% reduction in holding cost is consistent with Indian EOQ-based and

ABC-VED-based inventory studies, which report cost reductions of 20–35%. However, this study extends that literature in two respects: the **80.9% stockout reduction** is substantially larger than typically reported, plausibly because the baseline service level (83.1%) was unusually low, leaving greater headroom for improvement; and the **66.1% wastage reduction** is rarely quantified in Indian hospital pharmacy studies at all.

The finding that 70.8% of NDDS formulations exhibited non-smooth demand—significantly higher than the 41.2% observed among conventional formulations in the same hospitals—provides empirical support for treating NDDS as a distinct inventory category rather than aggregating it with general pharmaceuticals. This has direct formulary and procurement policy implications.

5.2 The Class-Matching Finding

The study's central methodological result is that LSTM and XGBoost were statistically indistinguishable in aggregate but significantly different within every demand class. This has two implications.

Theoretically, it supports the Syntetos–Boylan premise that demand-pattern characteristics, not dataset-level properties, determine method suitability (Syntetos et al., 2005). The mechanism is intuitive: tree ensembles partition feature space using dense lag structures, which suits smooth and erratic series; recurrent networks maintain a cell state across arbitrary intervals, which suits intermittent and lumpy series where the *timing* of the last demand, not just its magnitude, is informative.

Practically, it implies that hospitals need not select a single forecasting model. A class-matched ensemble requires only a pre-processing classification step and two model pipelines—a modest additional complexity for a 44.1% accuracy gain over the best uniform model.

5.3 Implementation Barriers

Quantitative gains notwithstanding, five barriers emerged that would constrain real-world deployment.

(i) Cold-chain handling capacity. The 39.6% increase in cold-chain handling events (96 → 134 per month) is a direct consequence of smaller, more frequent orders. At the time of study, cold storage at H1 was operating at 87% utilisation. More frequent handling increases both labour requirements and the risk of temperature excursions. This is a genuine counter-argument to EOQ-based order splitting and implies that cold-chain capacity, not inventory cost, may be the binding constraint in some hospitals. Administrators should assess handling capacity before adopting smaller order quantities for cold-chain NDDS.

(ii) Regulatory constraints on controlled substances. Transdermal fentanyl and buprenorphine

(2 of 24 formulations) fall under the Narcotic Drugs and Psychotropic Substances Act, 1985, which imposes separate licensing requirements and caps stocking at a maximum of three months' requirement. This ceiling limits EOQ flexibility upward but does not bind the downward direction, so optimization remained feasible—though the order-frequency increase was attenuated for these items.

(iii) Demand censoring. SHAP analysis revealed that prior-stockout indicators carried meaningful predictive weight (rank 9), implying that recorded demand partly reflects historical supply constraints rather than true patient need. ML models trained on censored demand will systematically under-forecast. This is a structural limitation of any retrospective forecasting study in a supply-constrained system and is discussed further in Section 5.5.

(iv) Pharmacist trust and algorithmic aversion. Five of 12 participants expressed initial scepticism about forecast-driven ordering. Framing the system as decision support rather than automated ordering, and allowing pharmacists to override recommendations with a logged rationale, was reported as essential to acceptance. This mirrors findings in the broader clinical decision support literature.

(v) Vendor lead-time variability. Lead-time CV across the nine primary suppliers was 0.41, with three suppliers accounting for 78% of late deliveries. Safety stock must absorb this variability, and no forecasting improvement can compensate for unreliable supply. Vendor performance management is therefore a necessary complement to, not a substitute for, demand analytics.

(vi) Model governance. No formal mechanism existed for updating models when the formulary changed or new NDDS formulations were introduced. New SKUs with fewer than 12 months of history cannot be reliably forecast using lag-based features and require either transfer learning from analogous formulations or a judgemental overlay.

5.4 Clinical Impact

Beyond cost, the intervention produced measurable availability improvements. The 38 avoided stockout episodes (47 → 9) affected an estimated 235 patient-episodes (mean 6.2 patients per episode). Of these, an estimated 28% would have required private retail purchase at a mean premium of approximately 42% over institutional procurement cost; at a mean NDDS course cost of ₹46,000, this represents avoided patient-borne expenditure of approximately ₹12.7 lakh over 12 months. Therapy interruptions requiring therapeutic substitution fell from 38 to 7 episodes, and delayed treatment initiations fell from 19 to 4.

These figures are estimates based on pharmacist-reported episode characteristics and should be

interpreted as indicative rather than precise. Nevertheless, they establish that the financial case is accompanied by a clinical case—an argument that resonates more strongly with clinical stakeholders than cost savings alone.

5.5 Limitations

Design. The study is retrospective and simulates rather than observes ML-assisted ordering. Real-world adherence to algorithmic recommendations, and the behavioural response of prescribers to improved availability, remain untested. A prospective pilot is the logical next step.

Demand censoring. As noted, recorded demand is partially supply-constrained, so the models may under-forecast true need. Approaches such as expectation-maximisation for censored demand or the use of prescription-initiation records as a demand proxy could address this in future work.

Generalisation. The three participating hospitals are private or trust-run tertiary facilities in a rapidly developing capital region. Findings may not transfer to government hospitals operating under APMSIDC procurement, where lead times, pricing, and formulary constraints differ substantially. Similarly, the Amaravati context—expanding capacity, nascent analytics—may limit transferability to mature hospital systems.

Scope. The NDDS portfolio comprised 24 formulations. Larger portfolios may exhibit additional demand archetypes not represented here. Prescriber behaviour was treated as an exogenous feature rather than modelled endogenously; availability-induced demand is therefore only partially captured.

Cost assumptions. Holding cost rate (22.5%), service level (95%), and baseline wastage (5.9%) are hospital-specific. Sensitivity analysis bounded the impact of these assumptions, but the absolute magnitude of savings is not directly portable to other settings.

6. Managerial and Policy Implications

For hospital administrators.

- Classify NDDS formulations by demand pattern before selecting a forecasting model. Deploy XGBoost for smooth and erratic series; LSTM for intermittent and lumpy series. Minimum data requirement: 24 months of dispensing records.
- Prioritise wastage reduction over holding-cost reduction in the business case; it accounts for approximately 59% of gross savings.
- Assess cold-chain handling capacity before adopting smaller, more frequent order quantities for cold-chain NDDS. Where capacity is binding, relax the EOQ constraint for those items.

- Budget for governance: model retraining on formulary change, analyst time, and an override mechanism with logged rationale.

For policymakers.

- Integrate forecasting modules into e-Aushadhi so that transactional data becomes forward-looking. The analytical layer, not the data layer, is the missing component.
- Permit EOQ-based order sizing for high-cost NDDS within state procurement rules; current rules favour fixed-cycle bulk ordering.
- Fund pharmacist training in forecast interpretation, not merely software deployment.
- Review NDPS stocking limits for controlled NDDS to permit algorithmic order sizing within existing caps.

For suppliers.

- Lead-time variability, not mean lead time, is the dominant driver of required safety stock. Sharing reliable lead-time distributions would allow hospitals to reduce safety stock without service-level loss.

7. Conclusion

This study demonstrates that demand-pattern-aware ML forecasting, integrated with continuous-review inventory optimization, can materially improve NDDS management in Indian tertiary care hospitals. A class-matched hybrid model reduced forecast error by 61.6% relative to incumbent practice, and the resulting inventory policy delivered simultaneous improvements in cost, wastage, stockouts, and service level—a Pareto improvement unattainable through inventory buffering alone.

The Amaravati case offers a replicable template for Tier-II Indian cities undergoing tertiary care expansion. The framework requires no proprietary software: open-source libraries and existing e-Aushadhi transaction data are sufficient. The principal obstacles are not algorithmic but infrastructural and behavioural—cold-chain handling capacity, regulatory stocking limits, demand censoring, and pharmacist trust.

Future research should prospectively validate the framework in live operations, model prescriber behaviour endogenously, apply censored-demand estimation, and explore federated learning architectures enabling multi-hospital model training without sharing patient-level data.

References

Ahmadi, E., Masel, D. T., Metcalf, A. Y., & Schuller, K. (2019). Inventory management of surgical supplies and sterile instruments in hospitals: A literature review. *Health Systems*, *8*(2), 134–151.

- Allen, T. M., & Cullis, P. R. (2013). Liposomal drug delivery systems: From concept to clinical applications. *Advanced Drug Delivery Reviews*, *65*(1), 36–48. *Confidence: High. Fits Section 2.2 — a widely cited review establishing the clinical and cost context of NDDS.*
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE. *Confidence: High. Fits Section 3.4 (Data Analysis) and Discussion.*
- Cameron, A., Ewen, M., Ross-Degnan, D., Ball, D., & Laing, R. (2009). Medicine prices, availability, and affordability in 36 developing and middle-income countries: A secondary analysis. *The Lancet*, *373*(9659), 240–249. *Confidence: High. Fits Section 2.3 — widely cited, provides context on medicine availability and pricing in India and comparable settings.*
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, *23*(3), 289–303. *Confidence: High. Fits Section 2.2.1 — the original paper on intermittent demand forecasting, foundational for your classification discussion.*
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs — Principles and practices. *Health Services Research*, *48*(6 Pt 2), 2134–2156. *Confidence: High. Fits Section 3.1 if you position this study within a broader mixed-methods programme.*
- Government of India. (1985). *The Narcotic Drugs and Psychotropic Substances Act, 1985* (Act No. 61 of 1985). Ministry of Finance, Department of Revenue.
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, *24*(2), 105–112. *Confidence: High. Fits Section 3.4 if you discuss content analysis alongside thematic analysis.*
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. *Confidence: High. Fits Section 3.3 (Sample Size Justification).*
- Hoffman, A. S. (2008). The origins and evolution of "controlled" drug delivery systems. *Journal of Controlled Release*, *132*(3), 153–163. *Confidence: High. Fits Section 2.1 — historical context for why NDDS behave differently from conventional drugs.*
- Kelle, P., Woosley, J., & Schneider, H. (2012). Pharmaceutical supply chain risks: A systematic review. *International Journal of Production Economics*, *135*(1), 312–319. *Confidence: Medium–High. Fits Section 2.3 — verify title and pages carefully.*
- Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*, *42*(8), 846–854. *Confidence: High. Fits Section 3.4 (Data Analysis) — a practical methods reference for medical education and health research.*
- Langer, R. (1998). Drug delivery and targeting. *Nature*, *392*(6679 Suppl), 5–10. *Confidence: High. Fits Section 2.1 — a foundational paper, highly cited across drug delivery literature.*
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE. *Confidence: Very high. Fits Section 3.5 (Trustworthiness) — the canonical source for credibility, transferability, dependability, and confirmability.*
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760. *Confidence: High. Fits Section 3.3 — justifies sample size using the "information power" concept rather than only saturation.*
- Management Sciences for Health. (2012). *MDS-3: Managing access to medicines and health technologies*. Management Sciences for Health. *Confidence: High. Fits Section 2.2 — standard reference for pharmaceutical management in low- and middle-income settings; a widely used practitioner reference.*
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE. *Confidence: Very high. Fits Section 3.4 — standard reference for coding and display of qualitative data.*
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, *16*(1), 1–13. *Confidence: High. Fits Section 3.5 (Trustworthiness) — practical criteria for demonstrating rigour in thematic analysis.*
- Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). SAGE.

Confidence: Very high. Fits Section 3 (Methodology)
— standard reference for qualitative design and interviewing.

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, *23*(4), 334–340.

Confidence: High. Fits Section 3.1 (Research Design)
— foundational for qualitative descriptive design.

Sandelowski, M. (2010). What's in a name? Qualitative description revisited. *Research in Nursing & Health*, *33*(1), 77–84.

Confidence: High. Fits Section 3.1 — companion piece to the 2000 paper.

Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and production management in supply chains* (4th ed.). CRC Press.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495–503.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357.

Torchilin, V. P. (2005). Recent advances with liposomes as pharmaceutical carriers. *Nature Reviews Drug Discovery*, *4*(2), 145–160.

Confidence: High. Fits Section 2.1 — establishes the science behind NDDS formulations.

Volland, J., Fügner, A., Schoenfelder, J., & Brunner, J. O. (2017). Material logistics in hospitals: A literature review. *Omega*, *69*, 82–101.

World Health Organization. (2021). *WHO model list of essential medicines* (22nd list). World Health Organization.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.

Confidence: Very high. Fits Section 3.1 — justifies the case study framing of the Amaravati context.