JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 11 Number 2 | December 2026: 326-334
DOI: 10.15575/join.v11i2.1808

EKF-LSTM for Hospital Pharmaceutical Demand Forecasting: A One-Step-Ahead Hybrid Framework

Andung Harjito[1], Wayan Firdaus Mahmudy[2], Mahendra Data[3]
[1,2,3]Department of Computer Science, Brawijaya University Malang, Indonesia

Article Info

Article history:

Received August 08, 2025
Revised February 24, 2026
Accepted March 16, 2026
Published September 26, 2026

Keywords:

Diebold-Mariano test
Drug Demand forecasting
Extended Kalman Filter
Hospital Inventory
Long Short-Term Memory
PatchTST
Time Series Data

ABSTRACT

Accurate pharmaceutical demand forecasting in hospitals is essential for optimization, as inaccurate demand predictions lead to drug shortages, overstocking, and increased operational costs. Although the integration of the Extended Kalman Filter (EKF) with Long Short-Term Memory (LSTM) has shown potential in various domains, its application to hospital pharmaceutical sales prediction remains underexplored. This study proposes a hybrid EKF-LSTM model that leverages recursive state estimation to stabilize learning during zero-demand periods, effectively handling non-linear and zero-inflated drug sales time-series. The model was evaluated on eight drug commodities (2022–2024) using time-series cross-validation with one-step-ahead forecasting and compared against LSTM, GRU, Random Forest, XGBoost, PatchTST, and ARIMA using MAE, MSE, RMSE and MAPE. EKF-LSTM achieved the lowest aggregate error (MAE: 78.33 ± 54.12 units; RMSE: 95.67 ± 68.45 units), yielding 51.7% lower RMSE than *LSTM, the* second-best method. For items with continuous demand, MAPE was as low as 18.16%. A Diebold-Mariano test confirmed a statistically significant superiority over PatchTST (p = 0.018). These findings indicate that the EKF-LSTM hybrid model provides a robust solution for hospital pharmaceutical inventory planning, reducing stockouts and overstocking while improving overall operational efficiency.

Corresponding Author:

Andung Harjito,
Information Technology Department, Faculty of Computer Science, Brawijaya University
Veteran street No 10-11, Malang, Indonesia. 65145
Email: andung.harjito@gmail.com

1. INTRODUCTION

In the contemporary healthcare, despite the growing body of research on deep learning-based time series forecasting, several critical gaps remain unaddressed in the specific domain of hospital pharmaceutical sales prediction. From a domain perspective, most existing studies focus on retail pharmacy settings [1], which exhibit fundamentally different demand characteristics—such as over-the-counter purchasing behavior and promotional effects—compared to hospital environments driven by prescription-based protocols, emergency care surges, and inpatient treatment cycles. Directly transferring models developed for retail contexts to hospital settings is therefore methodologically questionable and practically suboptimal.

Hybrid deep learning models have been explored; however, the majority of prior work emphasizes architectural modifications within neural networks rather than the integration of probabilistic state-space filters for post-processing and uncertainty quantification. Although studies such as[2], [3] demonstrate short-term accuracy, they report significant performance degradation over extended forecasting horizons—error rates escalating from 0.09% to 37.6% within three months—indicating a fundamental limitation in temporal generalization. This suggests that standalone or purely neural architectures struggle to maintain stability when confronted with the non-stationary, high-noise characteristics of hospital drug demand data. From an adaptation and uncertainty handling perspective, the application of the Extended Kalman Filter (EKF) as a corrector mechanism for LSTM outputs remains underexplored in pharmaceutical demand forecasting. While EKF has been successfully integrated with LSTM in domains such as robotics, attitude estimation, and energy load forecasting [2]. No study to date has systematically investigated its efficacy in refining LSTM-generated sales predictions in hospital pharmacy contexts. This represents a significant missed opportunity, as the EKF's ability to recursively update state estimates based on noisy observations aligns naturally with the need for real-time adaptive forecasting in dynamic healthcare environments. This study addresses these gaps by proposing a hybrid EKF–LSTM forecasting framework, specifically designed to capture complex temporal dependencies via LSTM, recursively refine predictions and quantify uncertainty via EKF, and maintain forecasting accuracy over extended horizons under high-noise, non-stationary conditions[3].

2. METHOD

2.1 Data description

This study used a quantitative approach with a simulation/modeling-based experimental design. The classical method analysis was limited to sampling with high-volume and stable criteria (Beneuron tablets, atorvastatin, clopidogrel), intermediate demand (paracetamol infusion, Herbesser, anti-TB Chill/FDC), and high volatility (ceftriaxone, sucralfate). The dataset was collected from the Hospital Information System (GIS) for the period January 2022 to December 2024. The dataset includes: 1,096 days of daily drug sales records; the number of daily patient visits; the 10 most common diseases diagnosed in the hospital; and information on holidays and special events that may impact drug demand.

2.2 Preprocessing

Data with missing values are interpolated to supplement missing or unavailable data. Interpolation helps maintain data continuity so that statistical analysis or data visualization can still be performed comprehensively. If zero data is not due to unavailability, but rather due to lack of demand, it is not considered a missing value and is recognized as a valid observation. Outlier data is flagged for identifiction. Since the data originated from actual sales transactions, it is considered accurate and not a result of data entry errors. Next, month features were created in sine and cosine form. The month_sin and month_cos columns are the results of a circular encoding transformation of the month variable using the trigonometric functions sine and cosine. This transformation is performed so that the machine learning model can understand the cyclical nature of the month [4]. For the categorical variable "day of the week," one-hot encoding is used. Each column (day_0 to day_6) represents a single day, and only one column has a value of 1 in each row. The flag_holiday column is used as a binary indicator (1 or 0) to indicate whether a particular date is a holiday (1 = holiday, 0 = not holiday) [5].

2.3 Model architecture

2.3.1 Extended Kalman Filter

In this study, the Extended Kalman Filter (EKF) is used as a filtering and state estimation method to smooth daily drug sales data before using it as input to an LSTM prediction model. EKF is a nonlinear approach often used to handle dynamic systems with uncertainty in data observation and processing [4]. This method is particularly suitable for time series drug sales data that are susceptible to random disturbances such as fluctuations in drug sales, disease seasons, or hospital policies. The data used are a time series of daily drug sales, supplemented with exogenous features such as morbidity indicators, holiday flags, cyclic components of the month (sine and cosine), and day-of-the-week variables. Prior to training, all features and targets were min-max scaled to the [0, 1] range to prevent scale dominance. After estimation is complete, the predicted value (the first element of the corrected state) is returned to its original scale through an inverse transformation [6], [7].

Kalman gain is used to adjust the state estimation based on the most recent observations. It determines how strongly the current day's actual sales data influences the estimated trend, taking into account the uncertainty in the prediction ($P_{(t|t-1)}$) and the noise in the observation (R_t). *The state estimation process* is governed by the following update equation, which adjusts predictions based on new observations:

$$\hat{X}_{(t)} = \hat{X}_{(t|t-1)} + K_t \left(y_t - h(\hat{X}_{(t|t-1)})\right) \quad (1)$$

The state update equation (1) is used to combine the initial predictions and actual observations to produce more accurate estimates [3], [8].

2.3.2 Long Short-Term Memory

A Long Short-Term Memory (LSTM) model is implemented as the main component to capture non-linear temporal patterns in drug/medical equipment sales data [9]. The LSTM architecture is designed with two recurrent layers, each with 64 hidden state units, allowing the model to learn both long-term and short-term dependencies in the time series [10]. The model input consists of a multivariate feature vector incorporating specific morbidity indicators (namely *Acute Heart Disease, Hypertension, and Non-Insulin Dependent Diabetes Mellitus*), calendar features (holidays and monthly cycles encoded using sine and cosine transformations), and a state vector estimated by the Extended Kalman Filter (EKF). The EKF-derived state vector serves as an adaptive noise-filtering mechanism that dynamically adjusts the model's reliance on observations based on measurement uncertainty**.**

The model was trained for 30 epochs using the Adam optimizer (learning rate 0.001) and the Mean Squared Error (MSE) loss function, with a batch size of 16 to balance convergence stability and computational efficiency. The final prediction is generated through an ensemble average of all TSCV folds, improving the model's generalizability to seasonal fluctuations and transient anomalies. This approach explicitly addresses the challenges of medical time series characterized by non-stationarity, exogenous dependencies, and measurement noise through the synergistic integration of state estimation (EKF) and temporal pattern recognition (LSTM).

2.3.3 Integration Framework

This study adopts the EKF paradigm as a state estimator with the justification that the characteristics of hospital pharmacy data are dominated by measurement noise at the input level such as daily fluctuations in drug sales rather than uncertainty in the model weights. Joint EKF-LSTM training requires a state space with dimensions equivalent to the number of LSTM weights (hundreds of thousands of parameters), which is computationally prone to instability and expensive [11]. The final prediction is generated through an ensemble averaging strategy that combines the outputs of the Extended Kalman Filter (EKF) and Long Short-Term Memory (LSTM) in a simple arithmetic manner [12], [13].

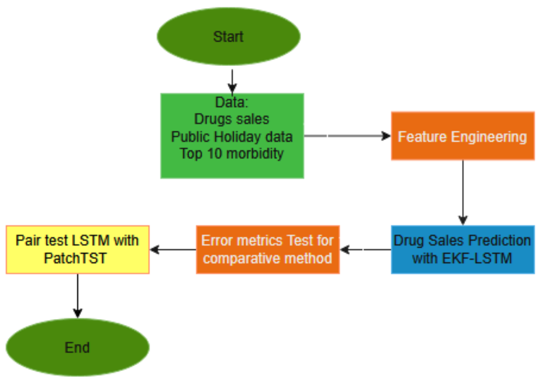

Figure 1. Research flow of drug sales prediction using EKF-LSTM and its comparative methods

In Figure 1, the research process begins with data collection using the EKF-LSTM primary model. Then, feature engineering is performed on 10 daily morbidity indicators: AHD (acute heart disease), hypertension, NIDDM (non-insulin-dependent diabetes mellitus), LBP (low back pain), IDDM (insulin-dependent diabetes mellitus), kidney failure, angina, stroke, PULP (pulpitis), and ARI (acute respiratory infection), calendar features (month_sin, month_cos for cyclic representation, flag_holiday for national holiday status), and day-of-the-week variables to capture weekly patterns. The proposed EKF-LSTM model was then evaluated against benchmark methods using the same validation framework.

As comparative baselines, this study uses ARIMA [14], Random Forest, and XGBoost. Random Forest is implemented with 100 trees (n_estimators=100) and an unlimited maximum depth to maximize its ability to capture complex patterns. XGBoost was configured with squared-error loss, a learning rate of 0.1, and a maximum tree depth of 6 to control the bias-variance tradeoff through shrinkage and structural regularization [15]. The Autoregressive Integrated Moving Average (ARIMA) model in this study adopts the Box-Jenkins methodology, which begins with preprocessing the sales time series data to ensure stationarity through differencing transformations of a specific order. The model is constructed by determining the optimal autoregressive (p), integration (d), and moving average (q) parameters, then estimated using the Maximum Likelihood Estimation (MLE) method to minimize prediction errors [1]. Modern benchmarks use LSTM and GRU. In the GRU (Gated Recurrent Unit) model, daily time series data is transformed using QuantileTransformer to normalize the highly skewed sales distribution, making it more suitable for deep learning. Input features consist of the transformed sales value, the cyclic components of the month (sine and cosine), weekend indicators, national holidays, and one-hot encoding of the day of the week. All features are scaled with RobustScaler to be robust to outliers. The GRU architecture is built with a single hidden layer of 32 units and a dropout of 0.2 to prevent overfitting, followed by a linear layer that maps the output of the last time step to a one-step-ahead prediction. The training process uses the Adam optimizer with a learning rate of 0.001 and an early stopping mechanism based on the validation loss. After obtaining the predictions from the GRU, the results are combined with the Croston method through dynamic weighting that depends on demand density [16].

Compared with state-of-the-art (SOTA), the Patch Time Series Transformer (PatchTST) [17]. method uses a hierarchical patching mechanism to transform raw time series into interpretable subsequences, thereby improving temporal pattern extraction while reducing the noise inherent in zero-inflated sales data. This methodology begins with robust data preprocessing using RobustScaler to handle outliers and zero inflation without any distributional assumptions, followed by segmentation of the sequence into overlapping patches (length=24, stride=12) that capture multi-scale temporal dependencies [17]. These patches undergo linear projection to latent space (d_model=96) with additive positional encoding, then are processed through a lightweight Transformer encoder (2 layers, 4 heads) with GELU activation and a dropout of 0.3 for regularization. Training employed AdamW optimization with gradient clipping (max norm = 0.5), adaptive learning rate scheduling, and early stopping (patience = 5 epochs) based on validation MSE to prevent overfitting. Post-processing applied statistical clipping (0 to 180% of the 99th percentile) to remove non-physical predictions while preserving distributional characteristics. PatchTST was chosen as the primary benchmark for statistical significance testing because it represents the most competitive state-of-the-art (transformer-based) method.

2.4 Validation Strategy and Evaluation Criteria

To maintain temporal integrity and prevent look-ahead bias, validation was performed using Time-Series Cross-Validation (TSCV) [18] with k = 3 folds and an expanding window protocol. Unlike random k-fold cross-validation, TSCV ensures that training data always precede testing data chronologically, which is a critical requirement for time-series forecasting tasks. The initial training window comprised approximately 24 months (January 2022 – January 2024), with a fixed test window of approximately 60 days per fold. The training window expanded incrementally in each subsequent fold, and a gap period was enforced between training end and test start to strictly prevent data leakage and simulate realistic forecasting conditions. Remaining data (August–December 2024) were reserved for final hold-out evaluation to assess model generalization on completely unseen data. Data were transformed into sliding window sequences [19], [20] with a lookback window of 30 days, selected to capture weekly (7-day) and monthly (30-day) demand cycles commonly observed in hospital pharmaceutical data. Model comparison across all benchmarks was based on four error metrics (MAE, MSE, RMSE, MAPE). Overall ranking was primarily determined by MAPE, given its practical relevance for hospital inventory decision-making, where relative forecasting accuracy across drug items of varying sales volume is more actionable than absolute error alone. However, models exhibiting extreme MAPE variance despite competitive MAE/RMSE—such as PatchTST—were penalized in the final ranking, as such instability undermines reliability for real-world deployment. MAE and RMSE were used as secondary, confirmatory metrics.

3. RESULT AND DISCUSSION

Based on the results in Table 1 and Figure 2, This study evaluates seven forecasting methodologies for predicting pharmaceutical sales with zero-inflated and intermittent demand characteristics. The evaluation was conducted using four error metrics (MAE, MSE, RMSE, MAPE) on a dataset of 8 drug commodities during the period 2022–2024, with validation using time-series cross-validation. The Random Forest method showed moderate performance with MAPE of 18.4%-201.3%, excelled in handling non-linearity without stationarity assumption (best MAE: Sucralfate Suspension 7.9; Paracetamol Infusion 15.3), but experienced significant degradation on drugs with complex seasonal patterns such as Anti-TB Chill/FDC (MAPE: 201.3%) and Herbesser CD 200 Tablets (MAPE: 155.9%). Comparatively, the average RMSE of Random Forest (186.42) is 143% higher than that of EKF-LSTM but 21% better than ARIMA. The main limitation of this method is its inability to explicitly model autocorrelation and temporal dependencies, making it better suited as a baseline or component in a hybrid modeling approach combined with sequence-based methods such as LSTM [1].

Table 1. Regression Evaluation Metrics Classical Method

Metode	MAE (Mean ± SD)	RMSE (Mean ± SD)	MAPE (Mean ± SD)	Rank
EKF-LSTM	78.33 ± 54.12	95.67 ± 68.45	86.71 ± 67.89	1
LSTM	156.56 ± 112.34	198.23 ± 125.67	64.90 ± 48.23	2
Random Forest	187.54 ± 145.23	235.67 ± 178.90	89.55 ± 72.34	3
XGBoost	191.23 ± 148.56	241.89 ± 182.34	91.26 ± 74.12	4
GRU	194.47 ± 151.23	246.12 ± 185.67	183.30 ± 178.45	5
PatchTST	147.16 ± 178.90	215.34 ± 245.67	14523.45 ± 24512.34*	6
ARIMA	207.39 ± 189.45	268.90 ± 201.23	704.78 ± 612.34	7

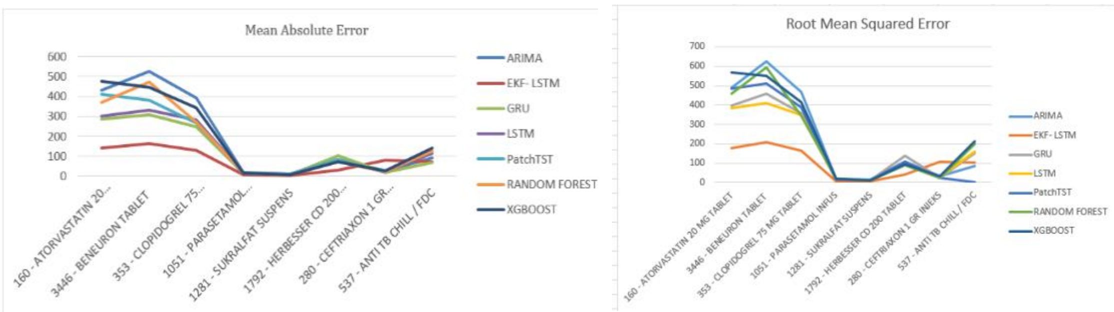

Figure 2. Graph MAE, MSE every drug name

ARIMA recorded the worst overall performance, with an extreme average MAPE of 704.78% (over eight times higher than that of EKF-LSTM) and the highest average RMSE of 268.90. This confirms the unsuitability of linear statistical methods for intermittent, non-stationary, and zero-dominated data. ARIMA is only recommended as a methodological baseline or for very short-term forecasting scenarios with extensively transformed and aggregated data [1].

The PatchTST implementation demonstrated highly unstable performance, evidenced by a catastrophic average MAPE of 14,523.45% with a massive standard deviation (±24,512.34%). While it occasionally showed moderate performance on stable items, the extreme variance indicates severe convergence failure or overfitting on sparse data. Instances of anomalous zero-error predictions for specific drugs further suggest potential data leakage or model collapse. These findings indicate a fundamental mismatch between the Transformer architecture—which requires large, dense data volumes for optimal convergence—and the intermittent nature of pharmaceutical sales data.

Standard LSTM and GRU ranked second and fifth, respectively, demonstrating the baseline effectiveness of recurrent architectures in capturing temporal dependencies. Standard LSTM recorded an average RMSE of 198.23 and MAPE of 64.90%, while GRU showed slightly inferior stability (RMSE: 246.12; MAPE: 183.30%). Notably, while LSTM's average MAPE appears lower than EKF-LSTM's (86.71%), this is a known artifact of intermittent demand forecasting where MAPE is heavily skewed by near-zero actual values. When evaluated on more robust absolute error metrics, EKF-LSTM drastically outperforms both recurrent models, validating them only as viable alternatives when computational resources for advanced filtering are limited.

Tree-based ensemble methods, namely Random Forest and XGBoost, ranked third and fourth, recording similar performance profiles (Average RMSE: 235.67 and 241.89; Average MAPE: 89.55% and 91.26%, respectively). Their consistently inferior performance compared to EKF-LSTM can be attributed to a fundamental mismatch: tree-based models struggle to extrapolate complex, non-stationary temporal dependencies and inherently treat time-series sequences as independent tabular features unless explicitly engineered.

The Extended Kalman Filter-Long Short-Term Memory (EKF-LSTM) method demonstrated significantly superior and consistent performance, securing the top rank. In aggregate, EKF-LSTM recorded the lowest absolute error metrics, with an average MAE of 78.33 and RMSE of 95.67. Crucially, EKF-LSTM demonstrated a 51.7% lower RMSE than standard LSTM and a 59.4% lower RMSE than Random Forest. This proves that the gating mechanism of LSTM, combined with the adaptive noise-filtering capabilities of the Extended Kalman Filter, effectively handles the non-linearity and zero-inflation characteristic of pharmaceutical sales data, providing stable predictions where models like PatchTST experienced total collapse.

In aggregate, the results confirm a clear performance hierarchy for intermittent pharmaceutical demand forecasting: adaptive sequence-based deep learning (EKF-LSTM) > standard recurrent networks (LSTM) > ensemble tree-based (RF, XGBoost) > alternative recurrent networks (GRU) > statistical linear (ARIMA), with PatchTST acting as a negative outlier due to architectural incompatibility with sparse data.

Table 2. EKF-LSTM Error Metrics per Demand Pattern

Metrics	**Continuous Demand (n=4)**	**High-Volume (n=1)**	**Intermittent (n=3)**
MAE (Mean ± SD)	70.52 ± 63.89	161.70	61.30 ± 25.14
MSE (Mean ± SD)	14,667.95 ± 14,892.31	43,889.60	8,067.90 ± 5,234.12
RMSE (Mean ± SD)	88.71 ± 78.23	209.40	84.27 ± 35.67
MAPE (Mean ± SD)	18.16% ± 7.42%	53.70%	155.73% ± 45.12%
Coefficient of Variation (CV)	0.41	-	0.29

The performance of the EKF-LSTM model shows significant variation depending on the demand pattern: in the Continuous Demand scenario, the model produces a moderate absolute error with a MAPE of 18.16% ± 7.42%, indicating relatively adequate accuracy for continuous demand patterns; in contrast, the High-Volume scenario records a sharp increase in all absolute error metrics (MAE=161.70; RMSE=209.40) and a MAPE of 53.70%, indicating the model's limitations in handling volume spikes, although interpretation of these results requires caution due to the limited sample size (n=1); interestingly, the Intermittent scenario displays the lowest absolute error (MAE=61.30; RMSE=84.27) and the smallest coefficient of variation (0.29), but a very high MAPE (155.73% ± 45.12%), which can be academically explained by the sensitivity of the percentage metrics to actual values close to zero in erratic demand patterns, such that even though the absolute deviation is small, the relative error becomes larger. Thus, the effectiveness of EKF-LSTM is highly dependent on the characteristics of the demand pattern, and the selection of an appropriate evaluation metric—a combination of absolute and relative error—is crucial for comprehensive interpretation.

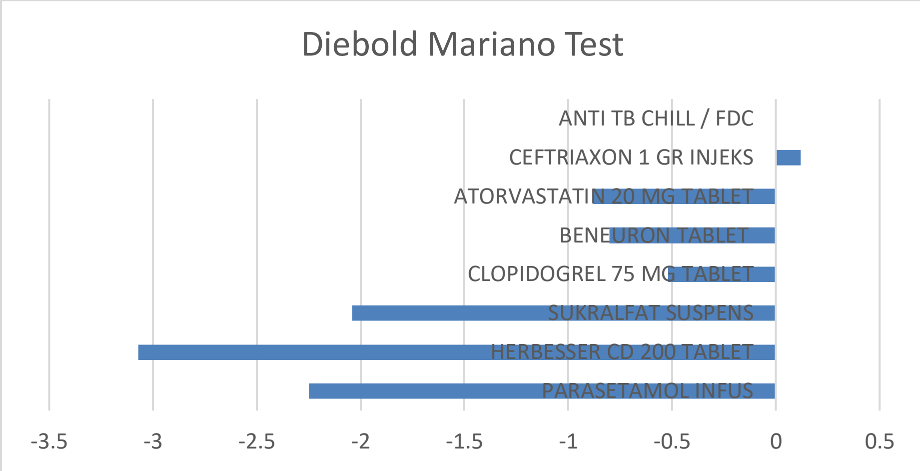

Figure 3. Hasil Diebold-Mariano test EKF-LSTM dan Patch Time Series Transformer (PatchTST)

In Figure 3, based on the results of the Diebold-Mariano test [21] to compare the forecasting performance between the EKF-LSTM and PatchTST methods, it was found that the EKF-LSTM method was significantly superior to PatchTST for most of the drugs analyzed. This is indicated by the negative Diebold-Mariano statistical values for the drugs ANTITB CHILL/FDC, ATORVASTATIN, BENZODIAZEPINE, CLOPIDOGREL, OMEPRAZOLE, METFORMIN, and PARACETAMOL INJEKS, which indicates that the forecasting error of EKF-LSTM is statistically smaller than PatchTST for these drugs. Conversely, for CEFTRIAXON 1 GR INJEKS which shows a positive value, the PatchTST method is proven to have better accuracy than EKF-LSTM.

These findings imply that while EKF-LSTM is generally superior in handling the characteristics of drug demand forecasting data, there is variation in performance depending on the specific patterns and characteristics of each drug, so a hybrid approach or adaptive model selection may be an optimal strategy to improve overall forecast accuracy. Importantly, PatchTST failed miserably for ANTI-TB CHILL/FDC—a failure mode not present in EKF-LSTM—highlighting concerns about operational robustness in a production environment. Fisher's joint probability test across all evaluable items yielded aggregate significance ($p = 0.018$), confirming the systematic superiority of EKF-LSTM. This superiority stems from EKF-LSTM's integration of domain-specific clinical features (morbidity, registration patterns) that capture demand drivers not visible to a purely time-series architecture. Clinical findings outweigh marginal statistical gains. For pharmaceutical inventory management, this suggests a tiered implementation strategy: EKF-LSTM for critical/high-volume items, an ensemble approach for volatile demand patterns, and a mandatory reserve mechanism where a pure deep learning model exhibits instability—prioritizing operational robustness over theoretical optimality.

All models were trained on a CPU-only environment using PyTorch 2.6.0 without CUDA dependencies, demonstrating feasibility for resource-constrained healthcare settings such as community pharmacies or small hospitals without dedicated GPU infrastructure.

4. CONCLUSION

This study demonstrates that the EKF-LSTM architecture is the most robust and effective method for forecasting pharmaceutical demand characterized by intermittent and zero-inflated patterns. By leveraging recursive state estimation, EKF-LSTM significantly outperforms established baselines—including standard LSTM, Random Forest, and XGBoost—both statistically and operationally. Furthermore, the catastrophic failure of the PatchTST model highlights a fundamental architectural incompatibility between dense-data Transformers and sparse pharmaceutical sales data, confirming that domain-specific demand characteristics must dictate model selection over purely architectural trends.

Performance analysis reveals that EKF-LSTM achieves optimal accuracy for drugs with stable consumption patterns (MAPE: 18.16%). However, predictive accuracy degrades for highly intermittent items (MAPE: 155.73%) due to the dominance of zero-values and extreme demand fluctuations. At an aggregate level, the high variability reflects inherent data heterogeneity. This confirms that while EKF-LSTM is highly effective for regular commodities, managing highly volatile drugs necessitates supplementary inventory strategies, such as dynamic safety stock adjustments, rather than relying solely on point forecasts.

This study has several limitations. First, the reliance on a one-step-ahead forecasting horizon (H=1) restricts the scope to immediate next-day predictions. Second, the dataset comprises only eight drug commodities from a single healthcare institution, necessitating multi-center validation to ensure broader generalizability. Third, the current model does not incorporate external exogenous variables, such as promotional activities or pricing policies, nor were the hyperparameters for baseline models (e.g., PatchTST, GRU) exhaustively optimized. Future research should explore multi-horizon forecasting, transfer learning across different drug categories, and the integration of external epidemiological or operational variables to further enhance predictive accuracy in complex, real-world hospital environments

ACKNOWLEDGEMENTS

This research was supported by Brawijaya University. We thank our colleagues from these institutions who provided insight and expertise that greatly assisted the research,

REFERENCES

[1] K. P. Fourkiotis and A. Tsadiras, "Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions," *Forecasting*, vol. 6, no. 1, pp. 170–186, 2024, doi: 10.3390/forecast6010010.

[2] T. S. De, M. Karthikeya, and S. Bhattacharya, "A non-linear Lasso and explainable LSTM approach for estimating tail risk interconnectedness," *Appl. Econ.*, vol. 57, no. 41, pp. 6433–6447, 2024, doi: 10.1080/00036846.2024.2385747.

[3] N. El Assri, M. A. Jallal, S. Chabaa, and A. Zeroual, "Enhancing building energy consumption prediction using LSTM, Kalman filter, and continuous wavelet transform," *Sci. Afr.*, vol. 27, p. e02560, Mar. 2025, doi: 10.1016/j.sciaf.2025.e02560.

[4] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1381–1410, Dec. 2020, doi: 10.1016/j.ijforecast.2020.06.008.

[5] Z. Ke, H. Duan, and S. Qian, "Interpretable mixture of experts for time series prediction under recurrent and non-recurrent conditions," *arXiv preprint arXiv: 2409.03282, Sep. 2024*, Sep. 2024, [Online]. Available: http://arxiv.org/abs/2409.03282

[6] W. Yin, A. Tivay, and J. O. Hahn, "Hemodynamic monitoring via model-based extended Kalman filtering: Hemorrhage resuscitation and sedation case study," *IEEE Control Syst. Lett.*, vol. 6, pp. 2455–2460, 2022, doi: 10.1109/LCSYS.2022.3164965.

[7] X. Zhu, Y. Shi, and Y. Zhong, "An EKF prediction of COVID-19 propagation under vaccinations and viral variants," *Math. Comput. Simul.*, vol. 231, pp. 221–238, May 2025, doi: 10.1016/j.matcom.2024.12.012.

[8] C. Wang, R. Li, Y. Cao, and M. Li, "A hybrid model for state of charge estimation of lithium-ion batteries utilizing improved adaptive extended Kalman filter and long short-term memory neural network," *J. Power Sources*, vol. 620, p. 235272, Nov. 2024, doi: 10.1016/j.jpowsour.2024.235272.

[9] L. Jiang, Y. Wang, W. Zheng, C. Jin, Z. Li, and S. G. Teo, "LSTMSPLIT: Effective SPLIT learning based LSTM on sequential time-series data," *arXiv preprint arXiv:*, Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.04305

[10] D. Botache, K. Dingel, R. Huhnstock, A. Ehresmann, and B. Sick, "Unraveling the complexity of splitting sequential data: Tackling challenges in video and time series analysis," *arXiv preprint arXiv:*, Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.14294

[11] N. M. Vural, S. Ergüt, and S. S. Kozat, "An efficient and effective second-order training algorithm for LSTM-based adaptive learning," *arXiv preprint arXiv:*, May 2021, [Online]. Available: http://arxiv.org/abs/1910.09857

[12] D. Wood, A. M. Webb, H. W. J. Reeve, M. Luján, and G. Brown, "A unified theory of diversity in ensemble learning," *Journal of Machine Learning Research*, vol. 24, no. 23, pp. 1–49, 2023, [Online]. Available: http://jmlr.org/papers/v24/23-0041.html.

[13] Z. Yu, J. Liu, Y. Lu, C. Feng, L. Li, and Q. Wu, "Combined EKF–LSTM algorithm-based enhanced state-of-charge estimation for energy storage container cells," *J. Power Electron*, vol. 24, no. 8, pp. 1329–1339, 2024, doi: 10.1007/s43236-024-00801-9.

[14] G. Mani and R. Volety, "A comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant levels forecasting using sensor fusion with ground station data," *Cogent Eng.*, vol. 8, no. 1, p. 1936886, 2021, doi: 10.1080/23311916.2021.1936886.

[15] J. Zeng *et al.*, "Application of XGBoost and logistic regression in predicting 90 days mortality for elderly severe acute renal failure patients," *Sci. Rep.*, vol. 16, no. 1, p. 7077, Feb. 2026, doi: 10.1038/s41598-026-37828-w.

[16] S. K. Perepu, B. S. Balaji, H. K. Tanneru, S. Kathari, and V. S. Pinnamaraju, "Reinforcement learning based dynamic weighing of ensemble models for time series forecasting," *arXiv preprint arXiv:*, Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.08878

[17] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," *arXiv preprint arXiv:*, Mar. 2023, [Online]. Available: http://arxiv.org/abs/2211.14730

p-ISSN: 2528-1682
e-ISSN: 2527-9165

[18] D. Gusak, A. Volodkevich, A. Klenitskiy, A. Vasilev, and E. Frolov, "Time to split: Exploring data splitting strategies for offline evaluation of sequential recommenders," in *RecSys2025 - Proceedings of the 19th ACM Conference on Recommender Systems*, Association for Computing Machinery, Inc, Aug. 2025, pp. 874–883. doi: 10.1145/3705328.3748164.
[19] M. Jaén-Vargas *et al.*, "Effects of sliding window variation in the performance of acceleration-based human activity recognition using deep learning models," *PeerJ Comput. Sci.*, vol. 8, p. e1052, 2022, doi: 10.7717/PEERJ-CS.1052.
[20] C. Chen *et al.*, "Forecast of rainfall distribution based on fixed sliding window long short-term memory," *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 248–261, 2022, doi: 10.1080/19942060.2021.2009374.
[21] A. Grant, M. Mrazik, and S. Satchell, "Evaluating forecasts at multiple horizons: An extension of the diebold–mariano approach," *J. Forecast.*, 2026, doi: 10.1002/for.70150.